us-09-540-234-1.rag

OM of: US-09-540-234-1 to: A_Geneseq_0601:*   out_format : pfs

Date: Sep 7, 2001 5:05 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09540234/runat_07092001_145854_19399/app_query.fasta_1.507
 -DB=A_Geneseq_0601 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH-0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN-2000000000
 -USER=US09540234_@CGN1_1_41 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
 Query: US-09-540-234-1
 Database: A_Geneseq_0601:*
 Query length: 444
 Database sequences: 412676
 Database length: 60623988
 Search time (sec): 65.090000

score_list:								
Sequence	Strd	Orig	ZScore	EScore	EScore	Len	!	Documentation
/SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:AAW11704	+	81.50	147.14	0.4108	637	!
/SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:AAW57224	+	81.50	146.68	0.4165	667	!
/SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:AAP90362	+	77.50	154.42	0.7793	132	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG58634	-	75.50	140.10	1.82	355	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG58633	-	75.50	139.33	1.86	383	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG58632	-	75.50	138.76	1.89	405	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY85569	-	75.00	119.76	3.68	2385	!
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90342	+	74.50	136.51	2.50	409	!
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW90341	+	74.50	134.70	2.64	489	!
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAW99484	+	73.50	147.09	2.27	116	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG29133	-	73.50	143.97	2.49	158	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG47928	+	73.50	143.97	2.49	158	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG29132	-	73.50	137.90	2.97	288	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG47927	+	73.50	137.90	2.97	288	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG29131	-	73.50	137.86	2.98	289	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG47926	+	73.50	137.86	2.98	289	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY85574	-	73.50	115.61	5.72	2608	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG06844	+	73.00	143.08	2.84	155	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG06843	+	73.00	139.31	3.18	225	!
/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAY01519	-	72.50	119.31	6.36	1462	!
/SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:AAR97246	+	72.50	108.00	8.86	4472	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG17227	+	71.50	132.93	5.28	307	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG17226	+	71.50	132.90	5.28	308	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG17225	+	71.50	132.83	5.29	310	!
/SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:AAR38470	-	71.50	107.20	11.23	3910	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR66462	+	71.50	107.05	11.28	3910	!
/SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:AAR52971	-	71.50	132.66	7.33	3969	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86251	+	70.00	132.31	7.40	237	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86252	+	70.00	132.31	7.40	237	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY79332	-	70.00	125.45	9.05	467	!
/SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:AAR13319	-	69.50	116.44	13.12	1023	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY75296	+	69.00	135.84	8.26	135	!
/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY75297	+	69.00	135.84	8.26	135	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86272	+	69.00	130.50	9.66	229	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86262	+	69.00	130.50	9.66	229	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86273	+	69.00	130.15	9.76	237	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86263	+	69.00	130.15	9.76	237	!
/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86264	+	69.00	130.15	9.76	237	!

/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV21873	+	69.00	119.43	13.38	684
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG03788	+	68.50	135.07	9.40	131
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB56069	-	68.50	126.82	11.98	296
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY76014	-	68.50	126.72	12.01	299
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY76077	-	68.50	126.72	12.01	299

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW11704

seq_documentation_block:
ID AAW11704 standard; Protein; 637 AA.
XX
AC AAW11704;
XX
DT 02-APR-1997 (first entry)
XX
DE High affinity Na+-dependent, L-Proline transporter.
XX
KW High-affinity sodium dependent L-Proline transporter; rat brain;
KW rTN2-2-20; neurotransmitter transporter; ss.
XX
OS Rattus rattus.
XX

FH	Key	Location/Qualifiers
FT	Domain	46..65
FT		/note= "Transmembrane domain 1"
FT	Domain	73..97
FT		/note= "Transmembrane domain 2"
FT	Domain	117..137
FT		/note= "Transmembrane domain 3"
FT	Domain	217..236
FT		/note= "Transmembrane domain 4"
FT	Domain	243..264
FT		/note= "Transmembrane domain 5"
FT	Domain	291..311
FT		/note= "Transmembrane domain 6"
FT	Domain	323..345
FT		/note= "Transmembrane domain 7"
FT	Domain	374..397
FT		/note= "Transmembrane domain 8"
FT	Domain	424..443
FT		/note= "Transmembrane domain 9"
FT	Domain	457..479
FT		/note= "Transmembrane domain 10"
FT	Domain	500..519
FT		/note= "Transmembrane domain 11"
FT	Domain	537..559
FT		/note= "Transmembrane domain 12"
FT	Modified-site	14..15
FT		/note= "Protein kinase C phosphorylation site"
FT	Modified-site	41..42
FT		/note= "cAMP-dependent protein kinase phosphorylation site"
FT	Modified-site	239..240
FT		/note= "Protein kinase C phosphorylation site"
FT	Modified-site	600..601
FT		/note= "Protein kinase C phosphorylation site"
FT	Domain	76..97
FT		/note= "Leucine zipper motif"
FT	Modified-site	182..183
FT		/note= "N-linked glycosylation site"
XX
PN	US5580775-A.
XX
PD	03-DEC-1996.
XX
PF	01-MAY-1992;	92US-0879617.
XX
PR	01-MAY-1992;	92US-0879617.
XX
PA	(UYEM-) UNIV EMORY.
XX
PI	Blakely RD,	Caron MG,	Fremeau RT;
XX

```
DR    WPI; 1997-033574/03.
DR    N-PSDB; AAT58853.
XX
PT    DNA encoding sodium dependent L-proline transporter protein - for
PT    expression of protein and for diagnostic, research and drug design
PT    use
XX
PS    Claim 8; Fig 2; 37pp; English.
XX
CC    This sequence represents a high-affinity sodium dependent L-Proline
CC    transporter derived from rat brain. This transporter protein has a
CC    mol. wt. of 71 kD.  Hydropathy analysis of the protein reveals the
CC    presence of 12 regions of significantly extended hydrophobicity
CC    suitable for the formation of transmembrane domains.  The N-terminus
CC    does not contain a readily identifiable signal sequence, suggesting
CC    that the N-terminus resides on the cytoplasmic face of the membrane.
CC    Comparison of the amino acid sequence with known neurotransmitter
CC    transporters reveals significant amino acid conservation,
CC    specifically 148 amino acids are conserved among cloned
CC    neurotransmitter transporters.  Many of the conserved residues are
CC    within or adjacent to the presumed transmembrane regions.
XX
SQ    Sequence    637 AA;

alignment_scores:
          Quality:        81.50               Length:       105
          Ratio:          1.455               Gaps:           2
Percent Similarity:      53.333        Percent Identity:    24.762

alignment_block:
US-09-540-234-1 x AAW11704         ..

Align seg 1/1    to: AAW11704    from: 1    to: 637

105 AATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAAACATGCCCTC  154
    :::||:||   :::||||||  :|||||||   ::|   ||||
174 AspGlyAsnGlyAlaLeuProLeuAsnLeuSerSerThrValSerProSe  190

155 GAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCG  204
    |::::::   |||||| :::         :::::|:::::::::|||:
190 rGluGluTyrTrpSerArgTyrValLeuHisIleGlnGlySerGlnGlyI  207

205 TCGATGTTATCGGAGCCATTCGTGCGTGCCTGTTTACTGTGTAGCCGCTTGATG  254
    ||  |||||                 ||||||
207 leGlyArgProGlyGluIleArg....................  214

255 ACTTCCTTGCATGCTACTCCGGTTGGAGCCATTTGAGTGCGTGTTTACCGTGC  304
    |||::                    :::::|  ||||||
215 ..............TrpAsnLeuCysLeuCysLeuLeuLeuAl  224

305 AGCACTCGAAGACTTCCTTGCGTGTACTCCGCTGGCAGCCATAGTCTGTTTGC  354
    | ::::::     |||||||:::::||| |||
224 aTrpValIleValPheLeuCysIleLeuLysGly................v  236

355 TGCCTGGTGCTGGGGACACCTGTACTACACAGTACTTCGTTTCCCTTTGTA  404
    :: ::: |||    |||||:::::::::::|||||||:::::|:::::
236 aLysSerSerGlyLysValValTyrPheThrAlaThrPheProTyrLeu  252

405 GTGGTACTACTTCTT  419
    ::::::|||:||||
253 IleLeuLeuMetLeu  257

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:AAW57224

seq_documentation_block:
ID    AAW57224 standard; Protein; 667 AA.
XX
AC    AAW57224;
XX
DT    04-AUG-1998  (first entry)
XX
DE    Rat proline transporter rPROT.
```

```
XX    Rat; proline transporter; rPROT; neurotransmission; CNS; L-proline;
KW    central nervous system; glutamatergic pathway; sodium.
KW
XX
OS    Rattus sp.
XX
FH    Key               Location/Qualifiers
FT    Misc-difference   1
FT                      /note= "encoded by TAT"
FT    Misc-difference   2
FT                      /note= "encoded by CTC"
FT    Misc-difference   3
FT                      /note= "encoded by AAA"
FT    Misc-difference   4
FT                      /note= "encoded by GGC"
FT    Misc-difference   5
FT                      /note= "encoded by GCA"
FT    Misc-difference   6
FT                      /note= "encoded by GAG"
FT    Misc-difference   7
FT                      /note= "encoded by ATA"
FT    Misc-difference   8
FT                      /note= "encoded by GGA"
FT    Misc-difference   9
FT                      /note= "encoded by CCA"
FT    Misc-difference   10
FT                      /note= "encoded by GTG"
FT    Misc-difference   11
FT                      /note= "encoded by CTC"
FT    Misc-difference   12
FT                      /note= "encoded by GGC"
FT    Misc-difference   13
FT                      /note= "encoded by GCC"
FT    Misc-difference   14
FT                      /note= "encoded by CGC"
FT    Misc-difference   15
FT                      /note= "encoded by TTG"
FT    Misc-difference   16
FT                      /note= "encoded by GCT"
FT    Misc-difference   17
FT                      /note= "encoded by GGC"
FT    Misc-difference   18
FT                      /note= "encoded by TGA"
FT    Misc-difference   19
FT                      /note= "encoded by CTG"
FT    Misc-difference   20
FT                      /note= "encoded by CGC"
FT    Misc-difference   21
FT                      /note= "encoded by TCT"
FT    Misc-difference   22
FT                      /note= "encoded by TGC"
FT    Misc-difference   23
FT                      /note= "encoded by AAG"
FT    Misc-difference   24
FT                      /note= "encoded by CAC"
FT    Misc-difference   25
FT                      /note= "encoded by CGG"
FT    Misc-difference   26
FT                      /note= "encoded by TGC"
FT    Misc-difference   27
FT                      /note= "encoded by CAG"
FT    Misc-difference   28
FT                      /note= "encoded by CTC"
FT    Misc-difference   29
FT                      /note= "encoded by TCC"
FT    Misc-difference   30
FT                      /note= "encoded by AAG"
FT    Region            44..45
FT                      /note= "protein kinase C phosphorylation site"
FT    Region            71..72
FT                      /note= "cAMP-dependant protein kinase phosphorylation
FT                      site"
FT    Domain            76..95
```

```
FT   Domain          /note= "membrane-spanning domain"
FT                   103..127
FT   Domain          /note= "membrane-spanning domain"
FT                   106..127
FT   Region          /note= "leucine zipper motif"
FT                   147..167
FT   Domain          /note= "membrane-spanning domain"
FT                   212..213
FT   Region          /note= "N-linked glycosylation site"
FT                   247..266
FT   Domain          /note= "membrane-spanning domain"
FT                   269..270
FT   Region          /note= "protein kinase C phosphorylation site"
FT                   272..294
FT   Domain          /note= "membrane-spanning domain"
FT                   321..341
FT   Domain          /note= "membrane-spanning domain"
FT                   353..375
FT   Domain          /note= "membrane-spanning domain"
FT                   404..427
FT   Domain          /note= "membrane-spanning domain"
FT                   454..473
FT   Domain          /note= "membrane-spanning domain"
FT                   487..509
FT   Domain          /note= "membrane-spanning domain"
FT                   530..549
FT   Domain          /note= "membrane-spanning domain"
FT                   567..589
FT   Region          /note= "protein kinase C phosphorylation site"
FT                   630..631
XX
PN   US5759788-A.
XX
PD   02-JUN-1998.
XX
PF   03-DEC-1996;    96US-0753985.
XX
PR   01-MAY-1992;    92US-0879617;
PR   03-DEC-1996;    96US-0753985.
XX
PA   (UYDU-) UNIV DUKE.
PA   (UYEM-) UNIV EMORY.
XX
PI   Blakely RD, Caron MG, Fremeau RT;
XX
DR   WPI; 1998-332127/29.
DR   N-PSDB; AAV28861.
XX
PT   L-proline transporter polypeptide  - used for study of the
PT   transporter and antibody against it for detection of the transporter
XX
PS   Claim 2; Column 29-34; 37pp; English.
XX
CC   The present sequence represents rat proline transporter (rPROT), which
CC   is a high-affinity, Na+-dependent L-proline transporter expressed in
CC   the brain. The present invention also describes an antibody that is
CC   specifically reactive with the L-proline transporter and does not
CC   cross-react with any other neurotransmitter transporter. The antibody
CC   can be used in assays for detecting the L-proline transporter in samples
CC   containing one or more proteins. The DNA encoding the polypeptide can
CC   be used for recombinant production of the transporter. The polypeptide
CC   can be used to make antibodies, in research studies, and in the design
CC   of modulating compounds for clinical treatments. This is useful as the
CC   transporter is thought to be involved in the release of amino acid
CC   transmitters for excitatory pathways in the central nervous system. The
CC   products can be used to study function and disorders involving these
CC   transporters.
XX
SQ   Sequence  667 AA;
XX
alignment_scores:
```

```
                    Quality:   81.50         Length:    105
                    Ratio:      1.455        Gaps:        2
Percent Similarity:  53.333   Percent Identity:  24.762

alignment_block:
US-09-540-234-1 x AAW57224

Align seg 1/1  to: AAW57224   from:  1  to: 667

105 AATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCAAACATGCCCTC 154
    :::||::::  ::||||||:::     ||||||   ::: |||||
204 AspGlyAsnGlyAlaLeuProLeuAsnLeuSerSerThrValSerProSe 220

155 GAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCG 204
    |::::: |   |||||   ::: :::::::::::::::::
220 rGluGluTyrTrpSerArgTyrValLeuHisIleGlnGlySerGlnGlyI 237

205 TCGATGTTATCGGAGCCATTCGTGCTGCTGTGTTACTGTGTAGCGCTTGATG 254
    ::             |||   |||||
237 leGlyArgProGlyGluIleArg................              244

255 ACTTCCTTGCTGTACTCCGCTTGGAGCCATTTGAGTGCTGCTGTTTACCGTGC 304
                            |||::   :::::||   ||
245 ..................TrpAsnLeuCysLeuCysLeuLeuLeuAl 254

305 AGCACTCGAAGACTTCCTTGCTGTACTCCGCTGGCAGCATAGTCTGTTTGC 354
    |    ::    ||||||:::: |:| |||      |||
254 aTrpValIleValPheLeuCysIleLeuLysGly..............V 266

355 TGCCTGCTCTGGGACACCTGCTGTACTACACAGTACTTCGTTTCCCTTTGTA 404
    :: ::: |||      ||||||:::::::|||||||::::::
266 aiLysSerSerGlyLysValValTyrPheThrAlaThrPheProTyrLeu 282

405 GTGGTACTACTTCTT 419
    ::::||||::|||
283 IleLeuLeuMetLeu 287
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:AAP90362

seq_documentation_block:
ID   AAP90362 standard; protein; 132.AA.
XX
AC   AAP90362;
XX
DT   01-NOV-1989  (first entry)
XX
DE   Recombinant Group A Eimeria tenella protein immunogen.
XX
KW   Recombinant proteins; Eimeria tenella; vaccines; poultry;
KW   coccidiosis.
XX
OS   Eimeria tenella.
XX
PN   EP324648-A.
XX
PD   19-JUL-1989.
XX
PF   13-JAN-1989;    89EP-0300329.
XX
PR   15-JAN-1988;    88US-0286934.
XX
PA   (MERI ) MERCK AND CO INC.
XX
PI   Liberator PA, Nollstadt KH, Turner MJ,  St John Crane M,  Karkhanis YD;
PI   Chakraborty PR, Profous-Juchelka H;
XX
DR   WPI; 1989-208595/29.
XX
PT   Recombinant Eimeria tenella proteins
PT   - used in vaccines to protect poultry against
PT   infection with Eimeria species which cause coccidiosis.
XX
PS   Claim 11; page 49; 58pp; English.
```

```
XX
CC   Group A Eimeria tenella protein immunogen, produced recombinantly.
CC   Used in vaccines to prevent coccidiosis from Eimeria tenella
CC   infection.
XX
SQ   Sequence   132 AA;

alignment_scores:
     Quality:    77.50         Length:    53
     Ratio:      2.583         Gaps:       1
Percent Similarity:  56.604    Percent Identity:  33.962

alignment_block:
US-09-540-234-1 x AAP90362      ..
Align seg 1/1  to: AAP90362   from:  1  to: 132

   194 CGCCAATGGCGCTGCGATGTTATCGGAGCCATTCGTCGTGCTGCTTACTGTGT 243
       |||::||||:||||||||||      |||  ::|:::::|||
    58 ArgSerTrpArgArgCysSerSerArgGlyAlaPheCysCy 74

   244 AGCGCTTGATGACTTCCTTGCTGTACTCCCGCTTGGAGCCATTGAGTGCT 293
       :::   :|||||   |||||  ||| ::: |||       :::||||
    74 sCysAspSerProPheLeuProCysPheAlaCysSerSerSerCysC 91

   294 GTTTACCGTGC...AGCACTCGAAGACTTCCTTGCTGCTGCTACTCGCTGGCAG 340
       ||   ||| |||||||||:::|||      ||:      :::
    91 ysCysCysAlaSerThrGluArgLysAsnCysPhePheAsnLeuThr 107

   341 CATAGTCTG 349
       |||::|||
   108 HisAlaLeu 110

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA2000.DAT:AAG58634

seq_documentation_block:
ID   AAG58634 standard; Protein; 355 AA.
XX
AC   AAG58634;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 75721.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;
PR   05-MAR-1999;    99US-0121825.
PR   09-MAR-1999;    99US-0123180.
PR   23-MAR-1999;    99US-0123548.
PR   25-MAR-1999;    99US-0125788.
PR   29-MAR-1999;    99US-0126264.
PR   01-APR-1999;    99US-0126785.
PR   06-APR-1999;    99US-0127462.
PR   08-APR-1999;    99US-0128234.
PR   16-APR-1999;    99US-0128714.
PR   19-APR-1999;    99US-0129845.
PR   21-APR-1999;    99US-0130077.
PR   23-APR-1999;    99US-0130449.
PR   23-APR-1999;    99US-0130510.
PR   28-APR-1999;    99US-0130891.
PR   30-APR-1999;    99US-0131449.
                     99US-0132048.
```

```
PR   30-APR-1999;    99US-0132407.
PR   04-MAY-1999;    99US-0132484.
PR   05-MAY-1999;    99US-0132485.
PR   06-MAY-1999;    99US-0132486.
PR   06-MAY-1999;    99US-0132487.
PR   07-MAY-1999;    99US-0132863.
PR   11-MAY-1999;    99US-0134256.
PR   14-MAY-1999;    99US-0134218.
PR   14-MAY-1999;    99US-0134219.
PR   14-MAY-1999;    99US-0134221.
PR   14-MAY-1999;    99US-0134370.
PR   18-MAY-1999;    99US-0134768.
PR   19-MAY-1999;    99US-0134941.
PR   20-MAY-1999;    99US-0135124.
PR   21-MAY-1999;    99US-0135353.
PR   24-MAY-1999;    99US-0135629.
PR   25-MAY-1999;    99US-0136021.
PR   27-MAY-1999;    99US-0136392.
PR   28-MAY-1999;    99US-0136782.
PR   01-JUN-1999;    99US-0137222.
PR   03-JUN-1999;    99US-0137528.
PR   04-JUN-1999;    99US-0137502.
PR   07-JUN-1999;    99US-0137724.
PR   08-JUN-1999;    99US-0138094.
PR   10-JUN-1999;    99US-0138540.
PR   10-JUN-1999;    99US-0138847.
PR   14-JUN-1999;    99US-0139119.
PR   16-JUN-1999;    99US-0139452.
PR   16-JUN-1999;    99US-0139453.
PR   17-JUN-1999;    99US-0139492.
PR   18-JUN-1999;    99US-0139454.
PR   18-JUN-1999;    99US-0139455.
PR   18-JUN-1999;    99US-0139456.
PR   18-JUN-1999;    99US-0139457.
PR   18-JUN-1999;    99US-0139458.
PR   18-JUN-1999;    99US-0139459.
PR   18-JUN-1999;    99US-0139460.
PR   18-JUN-1999;    99US-0139461.
PR   18-JUN-1999;    99US-0139462.
PR   18-JUN-1999;    99US-0139463.
PR   18-JUN-1999;    99US-0139750.
PR   18-JUN-1999;    99US-0139763.
PR   21-JUN-1999;    99US-0139817.
PR   22-JUN-1999;    99US-0139899.
PR   23-JUN-1999;    99US-0140353.
PR   23-JUN-1999;    99US-0140354.
PR   24-JUN-1999;    99US-0140695.
PR   28-JUN-1999;    99US-0140823.
PR   29-JUN-1999;    99US-0140991.
PR   30-JUN-1999;    99US-0141287.
PR   01-JUL-1999;    99US-0141842.
PR   02-JUL-1999;    99US-0142154.
PR   06-JUL-1999;    99US-0142055.
PR   08-JUL-1999;    99US-0142390.
PR   09-JUL-1999;    99US-0142803.
PR   12-JUL-1999;    99US-0142920.
PR   13-JUL-1999;    99US-0142977.
PR   14-JUL-1999;    99US-0143542.
PR   15-JUL-1999;    99US-0143624.
PR   16-JUL-1999;    99US-0144005.
PR   16-JUL-1999;    99US-0144085.
PR   19-JUL-1999;    99US-0144086.
PR   19-JUL-1999;    99US-0144325.
PR   19-JUL-1999;    99US-0144331.
PR   19-JUL-1999;    99US-0144332.
PR   19-JUL-1999;    99US-0144333.
PR   19-JUL-1999;    99US-0144334.
PR   20-JUL-1999;    99US-0144335.
PR   20-JUL-1999;    99US-0144352.
PR   20-JUL-1999;    99US-0144632.
PR   21-JUL-1999;    99US-0144884.
PR   21-JUL-1999;    99US-0144814.
PR   21-JUL-1999;    99US-0145086.
```

Mon Sep 10 07:42:16 2001

us-09-540-234-1.rag

Left column:

```
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301439.
XX
PR  25-FEB-1999; 99US-0121825.
PR  05-MAR-1999; 99US-0123180.
PR  09-MAR-1999; 99US-0123548.
PR  23-MAR-1999; 99US-0125788.
PR  25-MAR-1999; 99US-0126264.
PR  29-MAR-1999; 99US-0126785.
PR  01-APR-1999; 99US-0127462.
PR  06-APR-1999; 99US-0128234.
PR  08-APR-1999; 99US-0128714.
PR  16-APR-1999; 99US-0129845.
PR  19-APR-1999; 99US-0130077.
PR  21-APR-1999; 99US-0130449.
PR  23-APR-1999; 99US-0130510.
PR  23-APR-1999; 99US-0130891.
PR  28-APR-1999; 99US-0131449.
PR  30-APR-1999; 99US-0132048.
PR  30-APR-1999; 99US-0132407.
PR  04-MAY-1999; 99US-0132484.
PR  05-MAY-1999; 99US-0132485.
PR  06-MAY-1999; 99US-0132486.
PR  06-MAY-1999; 99US-0132487.
PR  07-MAY-1999; 99US-0132863.
PR  11-MAY-1999; 99US-0134256.
PR  14-MAY-1999; 99US-0134218.
PR  14-MAY-1999; 99US-0134219.
PR  14-MAY-1999; 99US-0134221.
PR  14-MAY-1999; 99US-0134370.
PR  18-MAY-1999; 99US-0134768.
PR  19-MAY-1999; 99US-0134941.
PR  20-MAY-1999; 99US-0135124.
PR  21-MAY-1999; 99US-0135353.
PR  24-MAY-1999; 99US-0135629.
PR  25-MAY-1999; 99US-0136021.
PR  27-MAY-1999; 99US-0136392.
PR  28-MAY-1999; 99US-0136782.
PR  01-JUN-1999; 99US-0137222.
PR  03-JUN-1999; 99US-0137528.
PR  04-JUN-1999; 99US-0137502.
PR  07-JUN-1999; 99US-0137724.
PR  08-JUN-1999; 99US-0138094.
PR  10-JUN-1999; 99US-0138540.
PR  10-JUN-1999; 99US-0138847.
PR  14-JUN-1999; 99US-0139119.
PR  16-JUN-1999; 99US-0139452.
PR  16-JUN-1999; 99US-0139453.
PR  17-JUN-1999; 99US-0139492.
PR  18-JUN-1999; 99US-0139454.
PR  18-JUN-1999; 99US-0139455.
PR  18-JUN-1999; 99US-0139460.
PR  18-JUN-1999; 99US-0139461.
PR  18-JUN-1999; 99US-0139462.
PR  18-JUN-1999; 99US-0139463.
PR  18-JUN-1999; 99US-0139457.
PR  18-JUN-1999; 99US-0139750.
PR  18-JUN-1999; 99US-0139763.
PR  21-JUN-1999; 99US-0139817.
PR  22-JUN-1999; 99US-0139899.
PR  23-JUN-1999; 99US-0140353.
PR  23-JUN-1999; 99US-0140354.
PR  24-JUN-1999; 99US-0140695.
PR  28-JUN-1999; 99US-0140823.
PR  29-JUN-1999; 99US-0140991.
PR  30-JUN-1999; 99US-0141287.
```

Right column:

```
PR  01-JUL-1999; 99US-0141842.
PR  01-JUL-1999; 99US-0142154.
PR  02-JUL-1999; 99US-0142055.
PR  06-JUL-1999; 99US-0142390.
PR  08-JUL-1999; 99US-0142803.
PR  09-JUL-1999; 99US-0142920.
PR  12-JUL-1999; 99US-0142977.
PR  13-JUL-1999; 99US-0143542.
PR  14-JUL-1999; 99US-0143624.
PR  15-JUL-1999; 99US-0144005.
PR  16-JUL-1999; 99US-0144085.
PR  16-JUL-1999; 99US-0144086.
PR  19-JUL-1999; 99US-0144325.
PR  19-JUL-1999; 99US-0144331.
PR  19-JUL-1999; 99US-0144332.
PR  19-JUL-1999; 99US-0144333.
PR  19-JUL-1999; 99US-0144334.
PR  19-JUL-1999; 99US-0144335.
PR  20-JUL-1999; 99US-0144352.
PR  20-JUL-1999; 99US-0144632.
PR  21-JUL-1999; 99US-0144884.
PR  21-JUL-1999; 99US-0144814.
PR  21-JUL-1999; 99US-0145086.
PR  22-JUL-1999; 99US-0145088.
PR  22-JUL-1999; 99US-0145085.
PR  22-JUL-1999; 99US-0145087.
PR  23-JUL-1999; 99US-0145089.
PR  23-JUL-1999; 99US-0145192.
PR  23-JUL-1999; 99US-0145145.
PR  26-JUL-1999; 99US-0145218.
PR  27-JUL-1999; 99US-0145224.
PR  27-JUL-1999; 99US-0145276.
PR  27-JUL-1999; 99US-0145913.
PR  28-JUL-1999; 99US-0145918.
PR  02-AUG-1999; 99US-0145919.
PR  02-AUG-1999; 99US-0145951.
PR  02-AUG-1999; 99US-0146386.
PR  03-AUG-1999; 99US-0146388.
PR  04-AUG-1999; 99US-0146389.
PR  05-AUG-1999; 99US-0147038.
PR  06-AUG-1999; 99US-0147204.
PR  06-AUG-1999; 99US-0147302.
PR  09-AUG-1999; 99US-0147192.
PR  10-AUG-1999; 99US-0147260.
PR  11-AUG-1999; 99US-0147303.
PR  12-AUG-1999; 99US-0147416.
PR  13-AUG-1999; 99US-0147493.
PR  13-AUG-1999; 99US-0147935.
PR  16-AUG-1999; 99US-0148171.
PR  17-AUG-1999; 99US-0148319.
PR  18-AUG-1999; 99US-0148341.
PR  18-AUG-1999; 99US-0148565.
PR  20-AUG-1999; 99US-0148684.
PR  20-AUG-1999; 99US-0149368.
PR  23-AUG-1999; 99US-0149175.
PR  23-AUG-1999; 99US-0149426.
PR  25-AUG-1999; 99US-0149902.
PR  26-AUG-1999; 99US-0149930.
PR  27-AUG-1999; 99US-0150566.
PR  27-AUG-1999; 99US-0150884.
PR  30-AUG-1999; 99US-0151065.
PR  31-AUG-1999; 99US-0151066.
PR  01-SEP-1999; 99US-0151080.
PR  07-SEP-1999; 99US-0151303.
PR  10-SEP-1999; 99US-0151438.
PR  13-SEP-1999; 99US-0151930.
PR  15-SEP-1999; 99US-0152363.
PR               99US-0153070.
PR               99US-0153758.
PR               99US-0154018.
```

```
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   14-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   22-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
```

```
                132  ::::::::::::|||:::                  |||||  ||||||    :::       ||
                     IleAlaAlaLeuAlaAlaGlyLeuAlaSerValThrSerAlaSerAsnCy 148

                144  TTTGAGACGACCCTCAAAAGGGCAGGGA 117
                     ||||||::::|||
                148  s..........LysGlySerGly 152


seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG58632

seq_documentation_block:
ID   AAG58632 standard; Protein; 405 AA.
XX
AC   AAG58632;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 75719.
XX
KW   Protein identification; signal transduction pathway; metabolic
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP103405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;  2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   07-MAY-1999;   99US-0132487.
PR   11-MAY-1999;   99US-0132863.
PR   14-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
```

```
alignment_scores:
      Quality:          75.50           Length:        110
      Ratio:             1.198           Gaps:            6
Percent Similarity:    57.273     Percent Identity:  27.273

alignment_block:
US-09-540-234-1/rev x AAG58633

Align seg 1/1  to: AAG58633  from: 1  to: 383

402  CAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCC................ 367
     |||||:::::  :::::::|||    |||  |||
 53  GlnGlnIleAsnGlnPheCysValGlyGlnAsnSerProProValLeuPh 69

366  .CCAGACCAGGCAGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA... 322
     |||||  ||||||||  |||     ::: :::||::::::::::::::
 69  eProAspAlaAlaAlaAspProLeuValAlaGlyLysIleMetAsnSerV 86

321  ..GGAAGTCTTCGAGTG...................CTGCCACGGTAAA 295
     |||::::  |||:::        |||    |||      ||||:::|||
 86  alGlyThrArgArgAlaGlyThrLeuSerLysTrpPheHisHisAsnLys 102

294  CAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCAAGCGC 245
     ::::|||||||:  :::::::  :::||||::::::::::::::::::
103  GluHisSerThrSerThrThrAsnLeuLysLysLysAspLysVa 119

244  TACCACAGTAAACAGCACACGAATGGCTCCGATAACAATCGACGCCATTGGC 195
     ;::||||||              ||| ::::|||:::::::
119  lArgValGluAsnAlaHisVal...........His.SerAlaValSer 131

194  GTTGTTAGTCTTGTACAAAACTACGGCTTCACCGACGTCTCGAGGGGCATG 145
```

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 14-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 18-OCT-1999; 99US-0159331.
PR 21-OCT-1999; 99US-0159637.
PR 21-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 22-OCT-1999; 99US-0160768.
PR 22-OCT-1999; 99US-0160770.
PR 22-OCT-1999; 99US-0160814.
PR 25-OCT-1999; 99US-0160815.
PR 25-OCT-1999; 99US-0160980.
PR 25-OCT-1999; 99US-0160981.
PR 26-OCT-1999; 99US-0160989.
PR 26-OCT-1999; 99US-0161404.
PR 26-OCT-1999; 99US-0161405.
PR 28-OCT-1999; 99US-0161406.
PR 28-OCT-1999; 99US-0161359.
PR 28-OCT-1999; 99US-0161360.
PR 29-OCT-1999; 99US-0161361.
PR 29-OCT-1999; 99US-0161920.
PR 29-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 75.50   Length: 110
Ratio: 1.198   Gaps: 6
Percent Similarity: 57.273   Percent Identity: 27.273

alignment_block:
US-09-540-234-1/rev x AAG58632   ..

Align seg 1/1 to: AAG58632 from: 1 to: 405

```
402 CAAAGGGAAACGAAGTACTGTAGTAGTACACAGGTGTCC.............. 367
    |||::::||| ||      :::::||||      ||      |||
 75 GlnGlnGlnPheCysValGlyGlnAsnSerProProValLeuPh 91

366 .CCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA.... 322
    |||||  |||||:|   ||   ::   :::|||||:::
 91 eProAspAlaAlaAspProLeuValAlaGlyLyLysIleMetAsnSerV 108

321 ..GGAAGTCTTCGAGTG..................CTGCACGGTAAA 295
    ||:::: |:::         ||      ||::::|||
108 aGlyThrArgArgAlaGlyThrSerLysTrpPheHisHisAsnLys 124

294 CAGCCACTCAAATGGCTCCAAGCGGAGTACAGCAGGAAGTCATCAAGCGC 245
    ::||||||||||    :::::  ::|||||||||      :::::::
125 GluHisSerThrSerSerThrThrAsnLeuLysLysLysAspLyssVa 141

244 TACACAGTAAACAGCACCACGAATGGCTCCGATAACATCGACGCCATTGGC 195
    :  ::||||||        ||| :::|||:::::::
141 lArgValGluAsnAlaHisVal.........His.SerAlaValSer 153

194 GTTGTTAGTCTTGTACAAACTACGGCTTCACCGAGCGTCTCGAGGGGCATG 145
    ::::::||||||||    ||||| |||||||      ||
154 IleAlaAlaLeuAlaAlaGlyLeuAlaServalThrSerAlaSerAsnCy 170

144 TTTGAGACGACCCTCAAAAGGGCAGGGA 117
    |      ||||||:::|||
170 s............LysGlySerGly 174
```

CC The invention provides vertebrate (human) protein homologue of a UNC-53
CC protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
CC their plus ends. The UNC-53 sequences are used to promote neural
CC regeneration, revascularization and wound healing; also for treating
CC neurodegenerative disease, acute traumatic injury, fibrotic disease and
CC autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
CC polynucleotides can be used for recombinant production of the proteins,
CC as a source of probes for detecting allelic variants and polymorphisms,
CC for sequencing genomic DNA and for detecting UNC-53 expression; and as
CC source of therapeutic antisense sequences. Cells that express the
CC protein are used to identify regulators of cell shape, growth, motility
CC and migration. They can also be used to identify proteins that are
CC involved in signal transduction pathways also involving UNC-53, and to
CC identify compounds that alter attachment of UNC-53 to microtubules. A
CC target gene coupled to a UNC-53 encoding sequence may be used to deliver
CC the target gene to a cellular microtubule or its plus ends. The present
CC sequence represents the amino acid sequence of the third human homologue
CC of UNC-53, designated hs-UNC-53/3.
XX
SQ Sequence   2385 AA;

alignment_scores:
     Quality: 75.00   Length: 140
     Ratio: 1.027   Gaps: 7
Percent Similarity: 52.143   Percent Identity: 30.714

alignment_block:
US-09-540-234-1/rev x AAY85569

Align seg 1/1 to: AAY85569 from: 1 to: 2385

```
415 AGTAGTACCACTACAAAGGGAAACGAAGTACTG.....TAGTACACAGG 372
    |||::||||||| ::||||||||          ||:::||
510 SerLysThrThrAlaAlaLysLysGluSerSerSerSerGl 526

371 TGTCCCAGCACCAGGCAGC.................AAACAGACTATGCTGC 337
    |::|||:||:|||||||||      |||||||||:::  |
526 yIleProLysProGlySerLysValProThrValLysGlnThrIleSerP 543

336 CAGCGAGTACAGCAAGGAGTCTTCGAGTGTGCCACGGTAAACAGCACTC 287
    ||::|||||||| |||:::||||    ||| |||
543 rOGlySerThrAlaSerLysGluSerGluLysPheArgThrThr...... 557

286 AAATGGCTCCAAGCGGAGTACAGCAAGGAGTCATCAAGCGCTACACAGT 237
    :::::|||||    |||:::::|||      |||:::      |||
558 .......LysGlySerProSerGlnSerLeuSerLysProIleThrMe 571

236 AAACAGCACGAATGGCTCCGATAACATCGACGCCATTGGCGTTGTT.A 188
    :::::  :::  ::|||      |||      |||      |||::
571 tGluLys........AlaSerAlaSerCysProAlaProLeuGluG 585

187 GTCTTGTACAAACTACGGCTTCACCGAGCGTCTCGAGGGGCATGTTTGAGA 138
    ::    ||||||||:::|||      |||:::|||
585 lyArgGluAlaGlyGlnAlaSerProSer.....GlySerCysThrMet 599

137 CGA....................CCCTCAAAAGGGCAGGGACTGTTTTCCAT 106
    |||          ||| |||
600 ThrValAlaGlnSerGlyGlnSerThrGlyAsnGlyAlaValGln.. 615

105 TATGCCGTGCCAGCTACCAGTAGCCGTGCCGTTGCCGGCTTGGAGTGTCTTTGTC 56
    :::||| |||  |||      |||:::::
```

```
  616 .LeuProGlnGlnGlnGlnHisSerHisProAsnThrAlaThrValAlaP 632                         :::|||::.  :::.   :::::   :::::   |||                              :::|||:.   :::   ::::::  :::::: |||||
                                                                                 43 LysHisGlnCysGlnGlnGlnArgGlnTyrThrGluSerAspLysArg.. 58

   55 CTTTCCTTCAGATGACCCAC 36                                                   387 TACTGTAGTACACAGGTGTCCCCAGACCAGGCAGCAAACAGACTATGCTG 338
      ||||||::.:  :::|||                                                            |||  :::::::|:|| :::  ::|:|||: :::  :::|
  632 roPheIleTyrArgAlaHis 638                                                   59 .ThrCysLeuGlnCysAspSerMetLysGlnArgGluArgGluLysGlnV 75

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAW90342                     337 CCAGCGAGTACAGCAAGGAA.......................... 318
                                                                                     :::|||:.  :::|||||:||
seq_documentation_block:                                                          75 aIGluGluGluThrArgGluLysGluGluHisGlnGluGlnHisGlu 91
ID  AAW90342 standard; protein; 409 AA.
XX                                                                               317 ...........................GTCTTCGAGTGCTGCACGGTAAACAG 292
AC  AAW90342;                                                                                                   ||||||| |                                      ||
XX                                                                                92 GluGluGluAspGluAsnProTyrValPheGluGluAspLysAspPheSe 108
DT  24-MAY-1999 (first entry)                                                                                   |||||||||
XX                                                                               291 CACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCAAGCGCTAC 242
DE  G. max truncated SBP2 protein.                                                    |||:::::|:   :::|||  ||| :::|||::|:|::|||::
XX                                                                               108 rThrArgValGluThrGluGly.GlySerIleArgValLeuLysLysPhe 124
KW  SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;
KW  seed; carbohydrate content; soybean.                                         241 ACAGTAAACAGCACACGAATGGCTCCGATAACATCGACGCCATTGGCGTT 192
XX                                                                                    ||| :::|||   ::::|||||  |||  :|||:::      ||||||
OS  Glycine max.                                                                 125 ThrGluLysSerLysLeuLeuGlnGlyIleGluAsnPheArgLeuAlaIl 141
XX
PN  WO9853086-A1.                                                               191 GTTAGTCTTGTACAAACTACGGCTTCAC........CGACGTC 157
XX                                                                                    :|||          |||  |||                  |||  ::::
PD  26-NOV-1998.                                                                 141 eLeu.........GluAlaArgAlaHisThrPheValSerProArgHisP 155
XX
PF  21-MAY-1998;   98WO-US10465.                                                156 TCGAGGGGCATGTTGAGACGACCCTCAAAAGGGCAGGGACTGTTTTCCA 107
XX                                                                                    :::::::|::||||                   |||||||||      :::
PR  22-MAY-1997;   97US-0047568.                                                155 heAspSerGluValValLeuPheAsnIleLysGlyArgAlaValLeuGly 171
XX
PA  (UNIW ) UNIV WASHINGTON STATE RES FOUND.                                    106 TTATGCCGTGCCAGCTACAGTAGCCCGTGTCCGGCCTTGGAGGTGTCTTGT 57
XX                                                                                    ||| |||    |||  ::::::::::::::::            ||||||
PI  Chao WS,  Grimes HD;                                                        172 LeuValArgGluSerGluThrGluLysIleThr.LeuGlu......... 184
XX
DR  WPI; 1999-070155/06.                                                         56 CCTTTCCTTCAGAT............GACCCACTAGGCGCTCCTGCATACATA 16
XX                                                                                    |||:::|||                  |||  |||:::|||    |||||:
PT  New modified plant sucrose binding proteins - used to develop               185 .....ProGlyAspMetIleHisIleProAlaGlyThrProLeuTyrIle 199
PT  transgenic plants which can have enhanced or decreased sucrose
PT  uptake activity in developing seeds                                         seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAW90341
XX
PS  Claim 7; Page 39-40; 58pp; English.                                         seq_documentation_block:
XX                                                                              ID  AAW90341 standard; protein; 489 AA.
CC  This sequence represents a novel sucrose binding protein, SBP2 isolated     XX
CC  from Glycine max. This protein is used in a method resulting in the         AC  AAW90341;
CC  production of a modified plant sucrose binding protein (SBP) which has a     XX
CC  modified amino acid sequence compared to a corresponding wild-type SBP,     DT  24-MAY-1999 (first entry)
CC  and where expression of the modified SBP in a yeast assay system confers    XX
CC  enhanced sucrose uptake compared to the corresponding wild-type SBP.        DE  G. max SBP2 protein.
CC  The products of the invention can be used for producing transgenic plants   XX
CC  which have modified sucrose uptake activity, particularly in developing     KW  SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;
CC  seeds. Enhanced sucrose uptake activity in developing seeds may be          KW  seed; carbohydrate content; soybean.
CC  desirable where it is an advantage to increase the carbohydrate content     XX
CC  of the seed (e.g. where the seed is the primary plant material harvested,   OS  Glycine max.
CC  such as soybean). In contrast, decreased sucrose uptake activity in         XX
CC  seeds might be desirable where the vegetative material of the plant is      PN  WO9853086-A1.
CC  harvested.  The SBP regulatory regions confer specific or enhanced          XX
CC  expression in developing seeds and so may be used to express any            PD  26-NOV-1998.
CC  transgene in developing seeds.                                              XX
XX                                                                              PF  21-MAY-1998;   98WO-US10465.
SQ  Sequence  409 AA;                                                           XX
                                                                                PR  22-MAY-1997;   97US-0047568.
alignment_scores:                                                               XX
     Quality:     74.50          Length:       168                             PA  (UNIW ) UNIV WASHINGTON STATE RES FOUND.
     Ratio:        0.887          Gaps:          8                             XX
Percent Similarity:  50.000   Percent Identity:  25.595                         PI  Chao WS,  Grimes HD;
                                                                                XX
alignment_block:                                                                DR  WPI; 1999-070155/06.
US-09-540-234-1/rev x AAW90342  ..                                              XX
                                                                                PT  New modified plant sucrose binding proteins - used to develop
Align seg 1/1  to: AAW90342  from: 1  to: 409                                   PT  transgenic plants which can have enhanced or decreased sucrose
                                                                                PT  uptake activity in developing seeds
  437 AGACACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGAAG 388                     XX
```

```
PS    Claim 13b; Page 37-38; 58pp; English.
XX
CC    This sequence represents a novel sucrose binding protein, SBP2 isolated
CC    from Glycine max. This protein is used in a method resulting in the
CC    production of a modified plant sucrose binding protein (SBP) which has a
CC    modified amino acid sequence compared to a corresponding wild-type SBP,
CC    and where expression of the modified SBP in a yeast assay system confers
CC    enhanced sucrose uptake compared to the corresponding wild-type SBP.
CC    The products of the invention can be used for producing transgenic plants
CC    which have modified sucrose uptake activity, particularly in developing
CC    seeds. Enhanced sucrose uptake activity in developing seeds may be
CC    desirable where it is an advantage to increase the carbohydrate content
CC    of the seed (e.g. where the seed is the primary plant material harvested,
CC    such as soybean). In contrast, decreased sucrose uptake activity in
CC    seeds might be desirable where the vegetative material of the plant is
CC    harvested. The SBP regulatory regions confer specific or enhanced
CC    expression in developing seeds and so may be used to express any
CC    transgene in developing seeds.
XX
SQ    Sequence   489 AA;

alignment_scores:
      Quality:    74.50            Length:   168
      Ratio:       0.887           Gaps:      8
Percent Similarity:  50.000     Percent Identity:  25.595

alignment_block:
US-09-540-234-1/rev x AAW90341     ..

Align seg 1/1  to: AAW90341  from: 1  to: 489

437 AGACAGGAGTACATGGCTAAGAGTAGTAGTACCACTACAAAGGGAAACGAAG 388
    ::|||:::: ::  ::::::  :::::: |||  ::  |||||
 43 LysHisGlnCysGlnGlnArgGlnTyrThrGluSerAspLysArg.. 58

387 TACTGTAGTACACAGGTGTCCCCAGACACCAGCCAGCCAAACAGACTATGCTG 338
    |||  ::::::|||  ::  ::|||: :: :::
 59 .ThrCysLeuGlnGlnCysAspSerMetLysGlnGluArgGluLysGlnV 75

337 CCAGCGAGTACAGCAAGGAA.................... 318
    ::::|||  :::::||| :::::
 75 aIGluGluGluThrArgGluLysGluGluGluHisGlnHisGlnHisGlu 91

317 ....................GTCTTCGAGTGCTGCACGGTAAACAG 292
                         |||||||  ::
 92 GluGluGluAspGluAsnProTyrValPheGluGluAspLysAspPheSe 108

291 CACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCAAGCGCTAC 242
    ||||::::    ::::||| ||| ::|||::||::|||::::::
108 rThrArgValGluThrGluGly.GlySerIleArgValLeuLysLysPhe 124

241 ACAGTAAACAGCACACGAATGGCTCCGATAAACATCGACGCCATTGGCGTT 192
     ||||||| :: ::::||  ||| :: :::::: |||||| ||
125 ThrGluLysSerLysLeuLeuGlnGlyIleGluAsnPheArgLeuAlaIl 141

191 GTTAGTCTTGTCAAACTACGGCTTCAC...............CGACGTC 157
    ::||  :: ::||  :||: ||| ||| ||:|::
141 eLeu.........GluAlaArgAlaHisThrPheValSerProArgHisP 155

156 TCGAGGGGCATGTTTGAGACGACCCTCAAAAGGGCAGGGACTGTTTTCCA 107
    :::::::::||| |||||||| :::
155 heAspSerGluValValLeuPheAsnIleLysGlyArgAlaValLeuGly 171

106 TTATGCCGTGCCAGCTACAGTCCGTGTCCCGGCTTGAGGTGTCTTTGT 57
    :::::::::|||||  ::::::|||::||:::|::::: ||||:|
172 LeuValArgGluSerGluThrGluLysIleThr.LeuGlu......... 184

 56 CCTTTCCTTCAGAT........GACCCACTAGGCGCTCCTGCATACATA 16
    ||:|| ||| |||  :::: ||| :|::||| |||||| ||||||
185 ....ProGlyAspMetIleHisIleProAlaGlyThrProLeuTyrIle 199
```

seg_documentation_block:
```
ID    AAW99484 standard; Protein; 116 AA.
XX
AC    AAW99484;
XX
DT    08-JUN-1999  (first entry)
XX
DE    Glycoprotein hormone alpha-subunit analogue alpha-Q27C.
XX
KW    Analogue; heterodimeric; glycoprotein hormone; hCG; hLH; hFSH; hTSH;
KW    human chorionic gonadotropin; human luteinising hormone; disulphide bond;
KW    human follicle stimulating hormone; human thyroid stimulating hormone;
KW    stability; primer; amplification; PCR; mutation.
XX
OS    Homo sapiens.
OS    Synthetic.
XX
PN    WO9858957-A2.
XX
PD    30-DEC-1998.
XX
PF    25-JUN-1998;    98WO-US13070.
XX
PR    25-JUN-1997;    97US-0050784.
XX
PA    (ISTF ) ARS APPLIED RES SYSTEMS HOLDING NV.
PA    (MCIN-) MCINNIS P G.
XX
PS    Example 12; Page 87; 139pp; English.
XX
PI    Moyle WR;
XX
DR    WPI; 1999-081219/07.
XX
PT    New stabilised glycoprotein hormones - particularly hCG, hLH, hFSH
PT    or hTSH, have an intersubunit disulphide crosslink between the
PT    alpha- and beta-subunits to improve stability
XX
CC    The invention relates to the production of analogues of a heterodimeric
CC    subunit glycoprotein hormone (GPH) e.g. human chorionic gonadotropin
CC    (hCG), human luteinising hormone (hLH), human follicle stimulating
CC    hormone (hFSH), human thyroid stimulating hormone (hTSH), and functional
CC    muteins, which are modified to contain an intersubunit disulphide bond,
CC    between an alpha-subunit cysteine and a beta-subunit cysteine, for
CC    improved stability, the analogue retaining at least a portion of the
CC    bioactivity for the corresponding native GPH receptor.  Sequences
CC    AAW99483-W99495 represent mutated alpha subunits used for the generation
CC    of the modified GPHs.  The improved analogues are designed specifically
CC    to reduce perturbation of the 3-dimensional structure of the hormone,
CC    thereby creating greater likelihood that the dimer will be formed in
CC    vivo and the formed dimer will have affinity for the native receptors and
CC    have agonistic activity on them.  The changes stabilise the GPHs and
CC    prolong the biological activities of the hormones.  The analogues can
CC    have uses as for the native GPHs.
XX
SQ    Sequence   116 AA;

alignment_scores:
      Quality:    73.50            Length:    89
      Ratio:       1.564           Gaps:      4
Percent Similarity:  52.809     Percent Identity:  25.843

alignment_block:
US-09-540-234-1 x AAW99484     ..

Align seg 1/1  to: AAW99484  from: 1  to: 116

145 CATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAAC 194
    ||||||       ::|::::: ::|||      :::::
 23 HisSerAlaProAspValGlnAspCysProGluCysThrLeuGlnGluAs 39
```

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA1999.DAT:AAW99484

```
195 GCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGCGTGTTACTGTGTA 244
     |||::  |||   :::   |||:::::|||||
 39 nProPhePheSerGlnProGlyAlaProIleLeuCysCys.......... 52

245 GCGCTTGATGACTTCCTTGCTGT...............ACTCCGCTT 276
     :::  ||||||              |||||||||
 53 ..........MetGlyCysPheSerArgAlaTyrProThrProLeu 65

277 GGAGCCATTGAGTGCGTGTTTACCGTGCGCAGCACTCGAAGACTTCCTTGCT 326
              :::  :::|||||  |||::
 66 ................ArgSerLysThrMetLeuVa 73

327 GTAC...TCGCTGGCAGCATAGTCTGTTTGCTGCCTGGTCTGGGGACACC 373
     ::::::::::  |||::|||||::::::      :::::
 73 lGlnLysAsnValThrSerGluSerThrCysCysValAlaLysSerTyrA 90

374 TGTGTACTACAGTACTT 390
     ::::|||||:::
 90 snArgValThrValMet 95
```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG29133

seq_documentation_block:
ID   AAG29133 standard; Protein; 158 AA.
XX
AC   AAG29133;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 34608.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.   PR   21-MAY-1999; 99US-0135353.
PR   05-MAR-1999; 99US-0123180.   PR   24-MAY-1999; 99US-0135629.
PR   09-MAR-1999; 99US-0123548.   PR   25-MAY-1999; 99US-0136021.
PR   23-MAR-1999; 99US-0125788.   PR   27-MAY-1999; 99US-0136392.
PR   25-MAR-1999; 99US-0126264.   PR   28-MAY-1999; 99US-0136782.
PR   29-MAR-1999; 99US-0126785.   PR   01-JUN-1999; 99US-0137222.
PR   01-APR-1999; 99US-0127462.   PR   03-JUN-1999; 99US-0137528.
PR   06-APR-1999; 99US-0128234.   PR   04-JUN-1999; 99US-0137502.
PR   08-APR-1999; 99US-0128714.   PR   07-JUN-1999; 99US-0137724.
PR   16-APR-1999; 99US-0129845.   PR   08-JUN-1999; 99US-0138094.
PR   19-APR-1999; 99US-0130077.   PR   10-JUN-1999; 99US-0138540.
PR   21-APR-1999; 99US-0130449.   PR   10-JUN-1999; 99US-0138847.
PR   23-APR-1999; 99US-0130510.   PR   14-JUN-1999; 99US-0139119.
PR   23-APR-1999; 99US-0130891.   PR   16-JUN-1999; 99US-0139452.
PR   28-APR-1999; 99US-0131449.   PR   16-JUN-1999; 99US-0139453.
PR   30-APR-1999; 99US-0132048.   PR   17-JUN-1999; 99US-0139492.
PR   30-APR-1999; 99US-0132407.   PR   18-JUN-1999; 99US-0139454.
PR   04-MAY-1999; 99US-0132484.   PR   18-JUN-1999; 99US-0139455.
PR   05-MAY-1999; 99US-0132485.   PR   18-JUN-1999; 99US-0139456.
PR   06-MAY-1999; 99US-0132486.   PR   18-JUN-1999; 99US-0139457.
PR   06-MAY-1999; 99US-0132487.   PR   18-JUN-1999; 99US-0139458.
PR   07-MAY-1999; 99US-0132863.   PR   18-JUN-1999; 99US-0139459.
PR   11-MAY-1999; 99US-0134256.   PR   18-JUN-1999; 99US-0139460.
PR   14-MAY-1999; 99US-0134218.   PR   18-JUN-1999; 99US-0139461.
PR   14-MAY-1999; 99US-0134219.   PR   18-JUN-1999; 99US-0139462.
PR   14-MAY-1999; 99US-0134221.   PR   18-JUN-1999; 99US-0139463.
PR   14-MAY-1999; 99US-0134370.   PR   18-JUN-1999; 99US-0139750.
PR   18-MAY-1999; 99US-0134768.   PR   18-JUN-1999; 99US-0139763.
PR   19-MAY-1999; 99US-0134941.   PR   21-JUN-1999; 99US-0139817.
PR   20-MAY-1999; 99US-0135124.   PR   22-JUN-1999; 99US-0139899.
                                      PR   23-JUN-1999; 99US-0140353.
                                      PR   23-JUN-1999; 99US-0140354.
                                      PR   24-JUN-1999; 99US-0140695.
                                      PR   28-JUN-1999; 99US-0140823.
                                      PR   29-JUN-1999; 99US-0140991.
                                      PR   30-JUN-1999; 99US-0141287.
                                      PR   01-JUL-1999; 99US-0141842.
                                      PR   02-JUL-1999; 99US-0142154.
                                      PR   02-JUL-1999; 99US-0142055.
                                      PR   06-JUL-1999; 99US-0142390.
                                      PR   08-JUL-1999; 99US-0142803.
                                      PR   09-JUL-1999; 99US-0142920.
                                      PR   12-JUL-1999; 99US-0142977.
                                      PR   13-JUL-1999; 99US-0143542.
                                      PR   14-JUL-1999; 99US-0143624.
                                      PR   15-JUL-1999; 99US-0144005.
                                      PR   16-JUL-1999; 99US-0144085.
                                      PR   16-JUL-1999; 99US-0144086.
                                      PR   19-JUL-1999; 99US-0144325.
                                      PR   19-JUL-1999; 99US-0144331.
                                      PR   19-JUL-1999; 99US-0144332.
                                      PR   19-JUL-1999; 99US-0144333.
                                      PR   19-JUL-1999; 99US-0144334.
                                      PR   19-JUL-1999; 99US-0144335.
                                      PR   20-JUL-1999; 99US-0144352.
                                      PR   20-JUL-1999; 99US-0144632.
                                      PR   21-JUL-1999; 99US-0144814.
                                      PR   21-JUL-1999; 99US-0144884.
                                      PR   21-JUL-1999; 99US-0145086.
                                      PR   22-JUL-1999; 99US-0145085.
                                      PR   22-JUL-1999; 99US-0145087.
                                      PR   22-JUL-1999; 99US-0145089.
                                      PR   22-JUL-1999; 99US-0145192.
                                      PR   23-JUL-1999; 99US-0145145.
                                      PR   23-JUL-1999; 99US-0145218.
                                      PR   23-JUL-1999; 99US-0145224.
                                      PR   26-JUL-1999; 99US-0145276.
                                      PR   27-JUL-1999; 99US-0145913.
                                      PR   27-JUL-1999; 99US-0145918.
                                      PR   27-JUL-1999; 99US-0145919.
                                      PR   28-JUL-1999; 99US-0145951.
                                      PR   02-AUG-1999; 99US-0146386.
```

```
PR    02-AUG-1999;    99US-0146388.
PR    02-AUG-1999;    99US-0146389.
PR    03-AUG-1999;    99US-0147038.
PR    04-AUG-1999;    99US-0147204.
PR    04-AUG-1999;    99US-0147302.
PR    05-AUG-1999;    99US-0147192.
PR    05-AUG-1999;    99US-0147260.
PR    06-AUG-1999;    99US-0147303.
PR    06-AUG-1999;    99US-0147416.
PR    09-AUG-1999;    99US-0147493.
PR    09-AUG-1999;    99US-0147935.
PR    10-AUG-1999;    99US-0148171.
PR    11-AUG-1999;    99US-0148319.
PR    12-AUG-1999;    99US-0148341.
PR    13-AUG-1999;    99US-0148565.
PR    13-AUG-1999;    99US-0148684.
PR    16-AUG-1999;    99US-0149368.
PR    17-AUG-1999;    99US-0149175.
PR    18-AUG-1999;    99US-0149426.
PR    20-AUG-1999;    99US-0149722.
PR    20-AUG-1999;    99US-0149723.
PR    20-AUG-1999;    99US-0149929.
PR    23-AUG-1999;    99US-0149902.
PR    23-AUG-1999;    99US-0149930.
PR    25-AUG-1999;    99US-0150566.
PR    26-AUG-1999;    99US-0150884.
PR    27-AUG-1999;    99US-0151065.
PR    27-AUG-1999;    99US-0151066.
PR    27-AUG-1999;    99US-0151080.
PR    30-AUG-1999;    99US-0151303.
PR    31-AUG-1999;    99US-0151438.
PR    01-SEP-1999;    99US-0151930.
PR    07-SEP-1999;    99US-0152363.
PR    10-SEP-1999;    99US-0153070.
PR    13-SEP-1999;    99US-0153758.
PR    15-SEP-1999;    99US-0154018.
PR    16-SEP-1999;    99US-0154039.
PR    20-SEP-1999;    99US-0154779.
PR    22-SEP-1999;    99US-0155139.
PR    23-SEP-1999;    99US-0155486.
PR    24-SEP-1999;    99US-0155659.
PR    28-SEP-1999;    99US-0156458.
PR    29-SEP-1999;    99US-0156596.
PR    04-OCT-1999;    99US-0157117.
PR    05-OCT-1999;    99US-0157753.
PR    06-OCT-1999;    99US-0157865.
PR    07-OCT-1999;    99US-0158029.
PR    08-OCT-1999;    99US-0158232.
PR    12-OCT-1999;    99US-0158369.
PR    13-OCT-1999;    99US-0159293.
PR    13-OCT-1999;    99US-0159294.
PR    13-OCT-1999;    99US-0159295.
PR    14-OCT-1999;    99US-0159329.
PR    14-OCT-1999;    99US-0159330.
PR    14-OCT-1999;    99US-0159331.
PR    14-OCT-1999;    99US-0159637.
PR    14-OCT-1999;    99US-0159638.
PR    18-OCT-1999;    99US-0159584.
PR    21-OCT-1999;    99US-0160741.
PR    21-OCT-1999;    99US-0160767.
PR    21-OCT-1999;    99US-0160768.
PR    21-OCT-1999;    99US-0160770.
PR    21-OCT-1999;    99US-0160814.
PR    21-OCT-1999;    99US-0160815.
PR    22-OCT-1999;    99US-0160980.
PR    22-OCT-1999;    99US-0160981.
PR    22-OCT-1999;    99US-0160989.
PR    25-OCT-1999;    99US-0161404.
PR    25-OCT-1999;    99US-0161405.
PR    25-OCT-1999;    99US-0161406.
PR    26-OCT-1999;    99US-0161359.
PR    26-OCT-1999;    99US-0161360.
PR    26-OCT-1999;    99US-0161361.
PR    28-OCT-1999;    99US-0161920.
PR    28-OCT-1999;    99US-0161992.
PR    28-OCT-1999;    99US-0161993.
PR    29-OCT-1999;    99US-0162142.
```

```
alignment_scores:
            Quality:     73.50             Length:          141
            Ratio:       1.131             Gaps:              6
Percent Similarity:      46.099   Percent Identity:  24.113

alignment_block:
US-09-540-234-1/rev x AAG29133  .

Align seg 1/1   to: AAG29133   from: 1  to: 158

  435 ACACGAGTACATGGCTAAGAAGTAGTAGTACCACTACAAAGG...........  397
      ::::||||:::::||||  :::::||||||  ::
   32 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa  48

  396 ..........GAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGG  357
                ||||  ||| ||||| :||||    :::|||
   48 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer..........  61

  356 CAGCAAACAGACTATGCTGCCAGCGAGTACAGCAGGAGAGTCTTCGAGTG  307
      ||| ::::::  ::::|||    :::|||   ||||||
   62 ..........CysLeuAspAsnLysLysGlnMetGluArgVal  73

  306 CTGCACGGTAAACAGCACTCAAATGGCTCCAGCGGGAGTACAGCAAGGAA  257
      :::::||||:::                     ||::
   74 PheTyrGlyGln...........                ArgGl  79

  256 GTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCGATAACATC  207
      |||||              |||||| :::||| ||||
   79 uGlnArgMetAlaThr..........HisGluValArgSerHisA  92

  206 GACCCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTC  157
      ||||||              :::  ||| :::::|||  ::::::
   92 rgArgGlu.......ProGluValArgAspAspGlyTrpMetGluIle  105

  156 TCGAGGGGCATGTTTGAGACG..........................  136
      ::: ||  ||||||||
  106 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysGlyI  122

  135 .ACCCTCAAAAGGCAGGGACTGTTTCCATTATGCCCGTGCCAGCTACAG  87
      ::::||           ::: :::|||||
  122 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI  139

   86 TAGCCGTGTCCGGCTTGGAGGTG  64
      ::||||:::::||:::||||||
  139 leAlaIleAspGlyIleGluVal  146
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG47928

seq_documentation_block:
ID    AAG47928 standard; Protein; 158 AA.
XX
AC    AAG47928;
XX
DT    18-OCT-2000  (first entry)
XX
DE    Arabidopsis thaliana protein fragment SEQ ID NO: 60463.
XX
KW    Protein identification; signal transduction pathway; metabolic pathway;
KW    hybridisation assay; genetic mapping; gene expression control; promoter;
XX    termination sequence.
XX
OS    Arabidopsis thaliana.
XX
PN    EP1033405-A2.
XX
PD    06-SEP-2000.
XX
PF    25-FEB-2000; 2000EP-0301439.
```

Code	Date	Reference		Code	Date	Reference
XX	25-FEB-1999;	99US-0121825.		PR	12-JUL-1999;	99US-0142977.
PR	05-MAR-1999;	99US-0123180.		PR	13-JUL-1999;	99US-0143542.
PR	09-MAR-1999;	99US-0123548.		PR	14-JUL-1999;	99US-0143624.
PR	23-MAR-1999;	99US-0125788.		PR	15-JUL-1999;	99US-0144005.
PR	25-MAR-1999;	99US-0126264.		PR	16-JUL-1999;	99US-0144085.
PR	29-MAR-1999;	99US-0126785.		PR	16-JUL-1999;	99US-0144086.
PR	01-APR-1999;	99US-0127462.		PR	19-JUL-1999;	99US-0144325.
PR	06-APR-1999;	99US-0128234.		PR	19-JUL-1999;	99US-0144331.
PR	08-APR-1999;	99US-0128714.		PR	19-JUL-1999;	99US-0144332.
PR	16-APR-1999;	99US-0129845.		PR	19-JUL-1999;	99US-0144333.
PR	19-APR-1999;	99US-0130077.		PR	19-JUL-1999;	99US-0144334.
PR	21-APR-1999;	99US-0130449.		PR	19-JUL-1999;	99US-0144335.
PR	23-APR-1999;	99US-0130510.		PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130891.		PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0131449.		PR	21-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0132048.		PR	21-JUL-1999;	99US-0144814.
PR	04-MAY-1999;	99US-0132407.		PR	22-JUL-1999;	99US-0145086.
PR	05-MAY-1999;	99US-0132484.		PR	22-JUL-1999;	99US-0145088.
PR	06-MAY-1999;	99US-0132485.		PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	99US-0132486.		PR	23-JUL-1999;	99US-0145087.
PR	07-MAY-1999;	99US-0132487.		PR	23-JUL-1999;	99US-0145089.
PR	11-MAY-1999;	99US-0132863.		PR	26-JUL-1999;	99US-0145192.
PR	14-MAY-1999;	99US-0134256.		PR	27-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134218.		PR	27-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134219.		PR	28-JUL-1999;	99US-0145224.
PR	18-MAY-1999;	99US-0134221.		PR	02-AUG-1999;	99US-0145276.
PR	19-MAY-1999;	99US-0134370.		PR	02-AUG-1999;	99US-0145913.
PR	20-MAY-1999;	99US-0134768.		PR	02-AUG-1999;	99US-0145918.
PR	21-MAY-1999;	99US-0134941.		PR	03-AUG-1999;	99US-0145919.
PR	24-MAY-1999;	99US-0135124.		PR	04-AUG-1999;	99US-0145951.
PR	25-MAY-1999;	99US-0135353.		PR	05-AUG-1999;	99US-0146386.
PR	27-MAY-1999;	99US-0135629.		PR	06-AUG-1999;	99US-0146388.
PR	28-MAY-1999;	99US-0136021.		PR	06-AUG-1999;	99US-0146389.
PR	01-JUN-1999;	99US-0136392.		PR	09-AUG-1999;	99US-0147038.
PR	03-JUN-1999;	99US-0136782.		PR	10-AUG-1999;	99US-0147204.
PR	04-JUN-1999;	99US-0137222.		PR	11-AUG-1999;	99US-0147302.
PR	07-JUN-1999;	99US-0137528.		PR	12-AUG-1999;	99US-0147192.
PR	08-JUN-1999;	99US-0137502.		PR	13-AUG-1999;	99US-0147260.
PR	10-JUN-1999;	99US-0137724.		PR	16-AUG-1999;	99US-0147303.
PR	10-JUN-1999;	99US-0138094.		PR	17-AUG-1999;	99US-0147416.
PR	14-JUN-1999;	99US-0138540.		PR	18-AUG-1999;	99US-0147493.
PR	16-JUN-1999;	99US-0138847.		PR	20-AUG-1999;	99US-0147935.
PR	16-JUN-1999;	99US-0139119.		PR	20-AUG-1999;	99US-0148171.
PR	17-JUN-1999;	99US-0139452.		PR	23-AUG-1999;	99US-0148319.
PR	18-JUN-1999;	99US-0139453.		PR	23-AUG-1999;	99US-0148341.
PR	18-JUN-1999;	99US-0139492.		PR	25-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.		PR	26-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.		PR	27-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.		PR	27-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.		PR	30-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.		PR	31-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.		PR	01-SEP-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.		PR	07-SEP-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.		PR	10-SEP-1999;	99US-0149902.
PR	21-JUN-1999;	99US-0139462.		PR	13-SEP-1999;	99US-0149930.
PR	22-JUN-1999;	99US-0139463.		PR	15-SEP-1999;	99US-0150566.
PR	23-JUN-1999;	99US-0139750.		PR	16-SEP-1999;	99US-0150084.
PR	23-JUN-1999;	99US-0139763.		PR	20-SEP-1999;	99US-0151065.
PR	24-JUN-1999;	99US-0139817.		PR	22-SEP-1999;	99US-0151066.
PR	28-JUN-1999;	99US-0139899.		PR	23-SEP-1999;	99US-0151080.
PR	29-JUN-1999;	99US-0140353.		PR	24-SEP-1999;	99US-0151303.
PR	30-JUN-1999;	99US-0140354.		PR	28-SEP-1999;	99US-0151438.
PR	01-JUL-1999;	99US-0140695.		PR		99US-0151930.
PR	02-JUL-1999;	99US-0140823.		PR		99US-0152363.
PR	06-JUL-1999;	99US-0140991.		PR		99US-0153070.
PR	08-JUL-1999;	99US-0141287.		PR		99US-0153758.
PR	09-JUL-1999;	99US-0141842.		PR		99US-0154018.
PR		99US-0142154.		PR		99US-0154039.
PR		99US-0142055.		PR		99US-0154779.
PR		99US-0142390.		PR		99US-0155139.
PR		99US-0142803.		PR		99US-0155486.
PR		99US-0142920.		PR		99US-0155559.
				PR		99US-0156458.

```
135 .ACCCTCAAAAGGGCAGGGACTGTTTCCATTATGCCGTGCCAGCTACAG  87
    :::::||              |||   :::: :
122 uValValMetSerLeuThrGluValGlyValLysGlyTyrGlnLeuLysGlyGlyIylGlyIyI 139

86 TAGCCGTGCCGGCTTGGAGGTG  64
   ::|||:::::|||:::||||||
139 leAlaIleAspGlyIyIleGluVal 146
```

seq_name: /SIDS8/gcgdata/geneseq/AA2000.DAT:AAG29132

seq_documentation_block:
```
ID   AAG29132 standard; Protein; 288 AA.
XX
AC   AAG29132;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 34607.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
```

```
PR   29-SEP-1999; 99US-0156596.
PR   04-OCT-1999; 99US-0157717.
PR   05-OCT-1999; 99US-0157753.
PR   06-OCT-1999; 99US-0157865.
PR   07-OCT-1999; 99US-0158029.
PR   08-OCT-1999; 99US-0158232.
PR   12-OCT-1999; 99US-0158369.
PR   13-OCT-1999; 99US-0159293.
PR   13-OCT-1999; 99US-0159294.
PR   13-OCT-1999; 99US-0159295.
PR   14-OCT-1999; 99US-0159329.
PR   14-OCT-1999; 99US-0159330.
PR   14-OCT-1999; 99US-0159331.
PR   14-OCT-1999; 99US-0159637.
PR   14-OCT-1999; 99US-0159638.
PR   18-OCT-1999; 99US-0159584.
PR   21-OCT-1999; 99US-0160741.
PR   21-OCT-1999; 99US-0160767.
PR   21-OCT-1999; 99US-0160768.
PR   21-OCT-1999; 99US-0160770.
PR   21-OCT-1999; 99US-0160814.
PR   21-OCT-1999; 99US-0160815.
PR   22-OCT-1999; 99US-0160980.
PR   22-OCT-1999; 99US-0160981.
PR   22-OCT-1999; 99US-0160989.
PR   25-OCT-1999; 99US-0161404.
PR   25-OCT-1999; 99US-0161405.
PR   25-OCT-1999; 99US-0161406.
PR   26-OCT-1999; 99US-0161359.
PR   26-OCT-1999; 99US-0161360.
PR   26-OCT-1999; 99US-0161361.
PR   28-OCT-1999; 99US-0161920.
PR   28-OCT-1999; 99US-0161992.
PR   28-OCT-1999; 99US-0161993.
PR   29-OCT-1999; 99US-0162142.
```

alignment_scores:
```
   Quality:  73.50                 Length:       141
     Ratio:   1.131                  Gaps:         6
Percent Similarity: 46.099.   Percent Identity: 24.113
```

alignment_block:
US-09-540-234-1/rev x AAG47928

Align seg 1/1 to: AAG47928 from: 1 to: 158

```
435 ACACGAGTACATGGCTAAGAGACTAGTAGTACCACTACAAAGG.......... 397
    :::||||:::||| :::::||||||||| ::
 32 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa  48

396 .........GAAACGAAGTACTGTAGTACACAGGTGTCCCCAGACCAGG  357
             ||| ||| ||||||  |:::||
 48 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer..........  61

356 CAGCAAACAGACTATGCTGCCAGCGAGTACAGCCAAGGAAGTCTTCGAGTG 307
    ||| :::::  :::|||      |||||
 62 .........CysLeuAspAsnLysLysGlnGlnMetGluArgVal  73

306 CTGCACGGTAAACAGCACTCAAATGGCTCCAGCGGAGTACAGCAAGGAA  257
    :::::||::::          |||::
 74 PheTyrGlyGln.................ArgGl  79

256 GTCATCAAGCGGCTACACAGTAAACAGCACACAGCAATGGCTCCGATAACATC 207
    ::: ||||||      |||||| ::||| ||||
 79 uGlnArgMetAlaThr.........HisGluValArgSerHisA  92

206 GACGGCCATTGGCCGTTGTTAGTCTTGTTACAAACTACGGCTTCACCGACGTC 157
    ||||:::   ::: ||||:: |||::: :::::
 92 rgArgGlu.........ProGluValArgAspAspAspGlyTrpMetGluIle 105

156 TCGAGGGGCCATGTTTGAGACG.......... 136
    ::: ||| ||||||||||
106 GluLeuGlyGlyPheGluThrGlySerGlyGluGlyAspAspAspLysGl 122
```

```
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   02-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   21-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   22-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   28-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.

PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   20-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   14-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   18-OCT-1999;   99US-0159637.
PR   21-OCT-1999;   99US-0159638.
PR   21-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   22-OCT-1999;   99US-0160770.
PR   22-OCT-1999;   99US-0160814.
PR   22-OCT-1999;   99US-0160815.
PR   25-OCT-1999;   99US-0160980.
PR   25-OCT-1999;   99US-0160981.
PR   25-OCT-1999;   99US-0160989.
PR   26-OCT-1999;   99US-0161404.
PR   26-OCT-1999;   99US-0161405.
PR   28-OCT-1999;   99US-0161406.
PR   28-OCT-1999;   99US-0161359.
PR   29-OCT-1999;   99US-0161360.
PR   29-OCT-1999;   99US-0161361.
PR   29-OCT-1999;   99US-0161920.
PR   29-OCT-1999;   99US-0161992.
PR   29-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.
```

```
alignment_scores:
                 Quality:   73.50              Length:  141
                   Ratio:   1.131                Gaps:    6
    Percent Similarity:   46.099   Percent Identity:  24.113

alignment_block:
```

US-09-540-234-1/rev x AAG29132

Align seg 1/1 to: AAG29132 from: 1 to: 288

```
435 ACACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGG............ 397
    ::||||::::::||| ::||||| :::
162 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 178

396 .......GAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGG 357
       ||| ||| ||||||| :::|||
178 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer......... 191

356 CAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTG 307
    ||| :::::: ::||| ::||| |||||
192 ...........CysLeuAspAsnLysLysGlnGlnMetGluArgVal 203

306 CTGCACGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAA 257
    :::||||::||| ||| |||
204 PheTyrGlyGln.....................ArgGl 209

256 GTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCGATAAACATC 207
    ::: ||||||| |||||| :::||| ||||
209 uGlnArgMetAlaThr...........HisGluValArgSerHisA 222

206 GACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTC 157
    |||||::: ::: ||||:: :::::|
222 rgArgGlu.........ProGluValArgAspAspGlyTrpMetGluIle 235

156 TCGAGGGGCATGTTTGAGACG.................. 136
    ::: ||| |||||||||
236 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysGl 252

135 .ACCCTCAAAAGGCAGGGACTGTTTTCCATTATGCCGTGCCAGCTACAG 87
    ::::::: ||| ||| ::: ::: :
252 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLeuGlyGlyI 269

86 TAGCCGTGTCCGGCTTGGAGGTG 64
   ::||||::::::||::||:||||
269 leAlaIleAspGlyIleGluVal 276
```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG47927

seq_documentation_block:
ID   AAG47927 standard; Protein; 288 AA.
XX
AC   AAG47927;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 60462.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   28-APR-1999; 99US-0130891.
PR   30-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   04-MAY-1999; 99US-0132407.
PR   05-MAY-1999; 99US-0132484.
PR   06-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   07-MAY-1999; 99US-0132487.
PR   11-MAY-1999; 99US-0132863.
PR   14-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   19-MAY-1999; 99US-0134221.
PR   20-MAY-1999; 99US-0134370.
PR   21-MAY-1999; 99US-0134768.
PR   24-MAY-1999; 99US-0134941.
PR   25-MAY-1999; 99US-0135124.
PR   27-MAY-1999; 99US-0135353.
PR   28-MAY-1999; 99US-0135629.
PR   01-JUN-1999; 99US-0136021.
PR   03-JUN-1999; 99US-0136392.
PR   04-JUN-1999; 99US-0136782.
PR   07-JUN-1999; 99US-0137222.
PR   08-JUN-1999; 99US-0137528.
PR   10-JUN-1999; 99US-0137502.
PR   10-JUN-1999; 99US-0137724.
PR   14-JUN-1999; 99US-0138094.
PR   16-JUN-1999; 99US-0138540.
PR   16-JUN-1999; 99US-0138847.
PR   17-JUN-1999; 99US-0139119.
PR   18-JUN-1999; 99US-0139452.
PR   18-JUN-1999; 99US-0139453.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   18-JUN-1999; 99US-0139763.
PR   21-JUN-1999; 99US-0139817.
PR   22-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.

PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   03-AUG-1999;   99US-0146389.
PR   04-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   05-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.

PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
                    Quality: 73.50   Length: 141
                    Ratio: 1.131   Gaps: 6
Percent Similarity: 46.099   Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG47927   .

Align seg 1/1 to: AAG47927 from: 1 to: 288

435 ACACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGG.......... 397
       :::||||:::::||| :::::|||||| :::
162 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 178

396 ........GAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGG 357
          ||| ||| ||||| :::|||
178 lGlyAsnGlyGluLysIleLysSerThrTyrLeuSer......... 191

356 CAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTG 307
        ||| |||||| :::||| |||||||
192 ....CysLeuAspAsnLysLysGlnMetGluArgVal 203

306 CTGCACGTAAACAGCACTCAAATGGCTCCAAGCGGGAGTACAGCAAGGAA 257
      ::::::|||::: |||::
204 PheTyrGlyGln.............................ArgGl 209

256 GTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCGATAAACATC 207
      :::: ||||| ||||| :::||| ||||
209 uGlnArgMetAlaThr...........HisGluValArgSerHisA 222

206 GACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTC 157
      |||||::   :: |||::: |||| :::||||||
222 rgArgGlu.........ProGluValArgAspAspGlyTrpMetGluIle 235

156 TCGAGGGGCATGTTTGAGACG......................... 136
       ::: ||| |||||||||
236 GluLeuGlyGluPheGluThrGlySerGlyGluGlyAspAspLysGl 252

135 .ACCCTCAAAAGGGCAGGGACTGTTTTCCATTATGCCGTGCCAGCTACAG 87
       :::::: ||| :::
252 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 269

 86 TAGCCGTGTCCGGCTTGGAGGTG 64
       :||:|:::||:::||:||||||
269 leAlaIleAspGlyIleGluVal 276

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG29131

```
seq_documentation_block:
ID   AAG29131 standard; Protein; 289 AA.
XX
AC   AAG29131;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 34606.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.         PR   18-JUN-1999; 99US-0139460.
PR   05-MAR-1999; 99US-0123180.         PR   18-JUN-1999; 99US-0139461.
PR   09-MAR-1999; 99US-0123548.         PR   18-JUN-1999; 99US-0139462.
PR   23-MAR-1999; 99US-0125788.         PR   18-JUN-1999; 99US-0139463.
PR   25-MAR-1999; 99US-0126264.         PR   18-JUN-1999; 99US-0139750.
PR   29-MAR-1999; 99US-0126785.         PR   18-JUN-1999; 99US-0139763.
PR   01-APR-1999; 99US-0127462.         PR   21-JUN-1999; 99US-0139817.
PR   06-APR-1999; 99US-0128234.         PR   22-JUN-1999; 99US-0139899.
PR   08-APR-1999; 99US-0128714.         PR   23-JUN-1999; 99US-0140353.
PR   16-APR-1999; 99US-0129845.         PR   23-JUN-1999; 99US-0140354.
PR   19-APR-1999; 99US-0130077.         PR   24-JUN-1999; 99US-0140695.
PR   21-APR-1999; 99US-0130449.         PR   28-JUN-1999; 99US-0140823.
PR   23-APR-1999; 99US-0130510.         PR   29-JUN-1999; 99US-0140991.
PR   23-APR-1999; 99US-0130891.         PR   30-JUN-1999; 99US-0141287.
PR   28-APR-1999; 99US-0131449.         PR   01-JUL-1999; 99US-0141842.
PR   30-APR-1999; 99US-0132048.         PR   01-JUL-1999; 99US-0142154.
PR   30-APR-1999; 99US-0132407.         PR   02-JUL-1999; 99US-0142055.
PR   04-MAY-1999; 99US-0132484.         PR   06-JUL-1999; 99US-0142390.
PR   05-MAY-1999; 99US-0132485.         PR   08-JUL-1999; 99US-0142803.
PR   06-MAY-1999; 99US-0132486.         PR   09-JUL-1999; 99US-0142920.
PR   06-MAY-1999; 99US-0132487.         PR   12-JUL-1999; 99US-0142977.
PR   07-MAY-1999; 99US-0132863.         PR   13-JUL-1999; 99US-0143542.
PR   11-MAY-1999; 99US-0134256.         PR   14-JUL-1999; 99US-0143624.
PR   14-MAY-1999; 99US-0134218.         PR   15-JUL-1999; 99US-0144005.
PR   14-MAY-1999; 99US-0134219.         PR   16-JUL-1999; 99US-0144085.
PR   14-MAY-1999; 99US-0134221.         PR   16-JUL-1999; 99US-0144086.
PR   14-MAY-1999; 99US-0134370.         PR   19-JUL-1999; 99US-0144325.
PR   18-MAY-1999; 99US-0134768.         PR   19-JUL-1999; 99US-0144331.
PR   19-MAY-1999; 99US-0134941.         PR   19-JUL-1999; 99US-0144332.
PR   20-MAY-1999; 99US-0135124.         PR   19-JUL-1999; 99US-0144333.
PR   21-MAY-1999; 99US-0135353.         PR   19-JUL-1999; 99US-0144334.
PR   24-MAY-1999; 99US-0135629.         PR   19-JUL-1999; 99US-0144335.
PR   25-MAY-1999; 99US-0136021.         PR   20-JUL-1999; 99US-0144352.
PR   27-MAY-1999; 99US-0136392.         PR   20-JUL-1999; 99US-0144632.
PR   28-MAY-1999; 99US-0136782.         PR   20-JUL-1999; 99US-0144884.
PR   01-JUN-1999; 99US-0137222.         PR   21-JUL-1999; 99US-0144814.
PR   03-JUN-1999; 99US-0137528.         PR   21-JUL-1999; 99US-0145086.
PR   04-JUN-1999; 99US-0137502.         PR   22-JUL-1999; 99US-0145088.
PR   07-JUN-1999; 99US-0137724.         PR   22-JUL-1999; 99US-0145085.
PR   08-JUN-1999; 99US-0138094.         PR   22-JUL-1999; 99US-0145087.
PR   10-JUN-1999; 99US-0138540.         PR   23-JUL-1999; 99US-0145089.
PR   10-JUN-1999; 99US-0138847.         PR   23-JUL-1999; 99US-0145192.
PR   14-JUN-1999; 99US-0139119.         PR   23-JUL-1999; 99US-0145145.
PR   16-JUN-1999; 99US-0139452.         PR   26-JUL-1999; 99US-0145218.
PR   16-JUN-1999; 99US-0139453.         PR   27-JUL-1999; 99US-0145224.
PR   17-JUN-1999; 99US-0139492.         PR   27-JUL-1999; 99US-0145276.
PR   18-JUN-1999; 99US-0139454.         PR   28-JUL-1999; 99US-0145913.
PR   18-JUN-1999; 99US-0139455.         PR   02-AUG-1999; 99US-0145918.
PR   18-JUN-1999; 99US-0139456.         PR   02-AUG-1999; 99US-0145919.
PR   18-JUN-1999; 99US-0139457.         PR   03-AUG-1999; 99US-0145951.
PR   18-JUN-1999; 99US-0139458.         PR   04-AUG-1999; 99US-0146386.
PR   18-JUN-1999; 99US-0139459.         PR   04-AUG-1999; 99US-0146388.
                                        PR   05-AUG-1999; 99US-0146389.
                                        PR   06-AUG-1999; 99US-0147038.
                                        PR   09-AUG-1999; 99US-0147204.
                                        PR   10-AUG-1999; 99US-0147302.
                                        PR   11-AUG-1999; 99US-0147260.
                                        PR   12-AUG-1999; 99US-0147303.
                                        PR   13-AUG-1999; 99US-0147416.
                                        PR   13-AUG-1999; 99US-0147493.
                                        PR   16-AUG-1999; 99US-0147935.
                                        PR   17-AUG-1999; 99US-0148171.
                                        PR   18-AUG-1999; 99US-0148319.
                                        PR   18-AUG-1999; 99US-0148341.
                                        PR   20-AUG-1999; 99US-0148565.
                                        PR   20-AUG-1999; 99US-0148684.
                                        PR                99US-0149368.
                                        PR                99US-0149175.
                                        PR                99US-0149426.
                                        PR                99US-0149722.
                                        PR                99US-0149723.
                                        PR                99US-0149929.
```

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

356 CAGCAAACAGACTATGTCGTGCCAGCGAGTACAGCCAAGGAAGTCTTCGAGTG 307
          ||| ||||||||| ::||| |||||
193 ....CysLeuAspAsnLysLysGlnGlnMetGluArgVal 204

306 CTGCACGTAAACAGCACTCAAATGGCTCCAAGCGCGAGTACAGCCAAGGAA 257
    ::::::::|||:::: |||::
205 PheTyrGlyGln...........................ArgGl 210

256 GTCATCAAGCGCTACACAGTAAACAGCACGAATGGCTCCGATAACATC 207
    ::: ||||||| ||||||| :::||| |||
210 uGlnArgMetAlaThr...........HisGluValArgSerHisA 223

206 GACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTC 157
    ||||||||| :::   :::||||:   ::||:||
223 rgArgGlu.........ProGluValArgAspAspGlyTrpMetGluIle 236

156 TCGAGGGGCATGTTGAGACG..................................... 136
    ::: ||| ||||||||||
237 GluLeuGlyGluPheGluThrGlyGlySerGlyGluGlyAspAspLysGl 253

135 .ACCCTCAAAAGGGCAGGGACTGTTTCCATTATGCCCGTGCCAGCTACAG 87
    :::::::|||||||||| ||| ::: |||
253 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 270

86 TAGCCGTGTCCGGCTTGGAGGTG 64
   ::||||:|||:::::||||||
270 leAlaIleAspGlyIleGluVal 277

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG47926

seq_documentation_block:
ID   AAG47926 standard; Protein; 289 AA.
XX
AC   AAG47926;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 60461.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   05-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0131449.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   05-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.

alignment_scores:
         Quality: 73.50   Length: 141
         Ratio: 1.131   Gaps: 6
Percent Similarity: 46.099   Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG29131   ..
Align seg 1/1 to: AAG29131 from: 1 to: 289

435 ACACGAGTACATGGCTAAGAAGTAGTACCACTACCAAAGG...........397
    ::|||:|||:::||| :::::|:|||| ::
163 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 179

396 .........GAAACGAAGTACTGTAGTACACAGGTGTCCCCAGACCAGG 357
    ||| ||||||||||||||||| ::|||
179 IGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer...........192

us-09-540-234-1.rag

```
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0134256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  18-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  16-JUN-1999;  99US-0139453.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  21-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.

PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148819.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  21-OCT-1999;  99US-0160815.
```

PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
        Quality: 73.50   Length: 141
        Ratio: 1.131   Gaps: 6
Percent Similarity: 46.099   Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x AAG47926   ..

Align seg 1/1 to: AAG47926 from: 1 to: 289

435 ACACGAGTACATGCCTAAGAAGTAGTACCACTACAAAGG............. 397
       ::||||::::||| || :::::||||| ::
163 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 179

396 .........GAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGG   357
             ||||| ||| :::||| ||||| ::|||
179 lGlyAsnGlyGluLysLysIleLysSerThrTyrLeuSer......... 192

356 CAGCAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTG 307
       ||| :::::|||:||||||||| |||||:|||| |||||
193 ....CysLeuAspAsnLysLysGlnGlnMetGluArgVal 204

306 CTGCACGGTAAACAGCACTCAAATGCTCCAGCGAGTACAGCAAGGAA 257
       :::||:||::: |||::
205 PheTyrGlyGln...........................ArgGl 210

256 GTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCGATAACATC 207
       ::: |||||| |||||| :::||| ::: ||
210 uGlnArgMetAlaThr...........HisGluValArgSerHisA 223

206 GACGGCCATTGGCGTTGTTAGTCTGTGTACAAACTACGGCTTCACCGACGTC 157
       ||||::   ::: ||||||| ::::::
223 rgArgGlu..........ProGluValArgAspAspTyrpMetGluIle 236

156 TCGAGGGGCATGTTTGAGACG.......................... 136
       ::: ||| |||||||||
237 GluLeuGlyPheGluThrGlySerGlyGluGlyGluAspAspLysGl 253

135 .ACCCTCAAAAGGGCAGGGGACTGTGTTTCCATTATGCCGTGCCAGCTACAG   87
       :::::|||| ||| ::: ::: |||
253 uValValMetSerLeuThrGluValLysGlyTyrGlnLeuLysGlyGlyI 270

86   TAGCCGTGTCCGGCTGGAGGTG   64
       ::|||||::::||::||||||
270 leAlaIleAspGlyIleGluVal 277

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY85574

seq_documentation_block:
ID   AAY85574 standard; Protein: 2608 AA.
XX
AC   AAY85574;
XX
DT   07-JUL-2000 (first entry)
XX
DE   Hs-UNC-53/3/GFP fusion insert of plasmid pGI3303.
XX
KW   UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
KW   anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
KW   antisclerotic; antimetastatic; anti-arthritic; autoimmune disease.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Misc-difference 1194
FT   /note= "unspecified"
XX
PN   WO9963080-A1.
XX
PD   09-DEC-1999.
XX
PF   02-JUN-1999;   99WO-EP03848.
XX
PR   03-JUN-1998;   98GB-0011962.
XX
PA   (JANC) JANSSEN PHARM NV.
XX
PI   Luyten WHML, De Raeymaeker MC, Geysen JJGH, Bogaert TAOE;
PI   Maerten LJS, Verhasselt P, Van De Craen M;
XX
DR   WPI; 2000-116370/10.
DR   N-PSDB; AAA07846.
XX
PT   Novel proteins and nucleic acids e.g. for treating neurodegeneration -
XX
PS   Disclosure; Fig 7e; 146pp; English.
XX
CC   The invention provides vertebrate (human) protein homologue of a UNC-53
CC   protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
CC   their plus ends. The UNC-53 sequences are used to promote neural
CC   regeneration, revascularization and wound healing; also for treating
CC   neurodegenerative disease, acute traumatic injury, fibrotic disease and
CC   autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
CC   polynucleotides can be used for recombinant production of the proteins,
CC   as a source of probes for detecting allelic variants and polymorphisms,
CC   for sequencing genomic DNA and for detecting UNC-53 expression; and as
CC   source of therapeutic antisense sequences. Cells that express the
CC   protein are used to identify regulators of cell shape, growth, motility
CC   and migration. They can also be used to identify proteins that are
CC   involved in signal transduction pathways also involving UNC-53, and to
CC   identify compounds that alter attachment of UNC-53 to microtubules. A
CC   target gene coupled to a cellular microtubule or its plus ends. The present
CC   sequence represents the amino acid sequence of a full-length Hs-UNC-53/3
CC   in fusion with GFP insert of plasmid pGI3305.
XX
SQ   Sequence 2608 AA;

alignment_scores:
        Quality: 75.00   Length: 140
        Ratio: 1.027   Gaps: 7
Percent Similarity: 52.143   Percent Identity: 30.714

alignment_block:
US-09-540-234-1/rev x AAY85574   ..

Align seg 1/1 to: AAY85574 from: 1 to: 2608

415 AGTAGTACCACTACAAAGGGAAACGAAGTACTG......TAGTACACAGG 372
       |||:||||||::   ::|||||| ::|| ::||
755 SerLysThrThrAlaAlaLysGluSerLeuIleProSerSerSerGl 771

371 TGTCCCAGACCAGGCAGC...............AAACAGACTATGCTGC 337
       |:::|||:|||||||||| |||||||||| |
771 yIleProLysProGlySerLysValProThrValLysGlnThrIleSerP 788

336 CAGCGAGTACAGCAAGGAAGTCTTCGAGTGTCGCACGGTAAACAGCACTC 287
       ||::||||||||| |||:||:::: ||| ::: |||
788 roGlySerThrAlaSerLysGluSerGluLysPheArgThrThr....... 802

.

```
 286 AAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCAAGCGGCTACACAGT 237
              :::::::: |||:::::::::: |||      |||::: |||::
 803 ........LysGlySerProSerGlnSerLeuSerLysProIleThrMe 816

 236 AAACAGCACACGAATGGCTCCGATAACATCGACGCCATTGGCGTTGTT.A 188
          ::::||        |||::|||   |||     |||        |
 816 tGluLys........AlaSerAlaSerCysProAlaProLeuGluG 830

 187 GTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGA 138
          ::    |||||||||||::::     |||:::|||
 830 lyArgGluAlaGlyGlnAlaSerProSer.....GlySerCysThrMet 844

 137 CGA..............CCCTCAAAAGGCAGGGACTGTTTTCCAT 106
                    |||  |||::||||    :::
 845 ThrValAlaGlnSerSerGlyGlnSerThrGlyAsnGlyAlaValGln.. 860

 105 TATGCGGTGCCAGTACAGTAGCCGTGTCCGGCTTGGAGGTGTCTTTGTC 56
          :::|||   |||   |||         |||            ::::|
 861 .LeuProGlnGlnGlnGlnHisSerHisProAsnThrAlaThrValAlaP 877

  55 CTTTCCTTCAGATGACCCAC 36
     ||||||::    ::|||
 877 roPheIleTyrArgAlaHis 883


seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG06844

seq_documentation_block:
ID   AAG06844 standard; Protein; 155 AA.
XX
AC   AAG06844;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3768.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;                          99US-0121825.
PR   05-MAR-1999;                          99US-0123180.
PR   09-MAR-1999;                          99US-0123548.
PR   23-MAR-1999;                          99US-0125788.
PR   25-MAR-1999;                          99US-0126264.
PR   29-MAR-1999;                          99US-0126785.
PR   01-APR-1999;                          99US-0127462.
PR   06-APR-1999;                          99US-0128234.
PR   08-APR-1999;                          99US-0128714.
PR   16-APR-1999;                          99US-0129845.
PR   19-APR-1999;                          99US-0130077.
PR   21-APR-1999;                          99US-0130449.
PR   23-APR-1999;                          99US-0130510.
PR   23-APR-1999;                          99US-0130891.
PR   28-APR-1999;                          99US-0131449.
PR   30-APR-1999;                          99US-0132048.
PR   30-APR-1999;                          99US-0132407.
PR   04-MAY-1999;                          99US-0132484.
PR   05-MAY-1999;                          99US-0132485.
PR   06-MAY-1999;                          99US-0132486.
PR   06-MAY-1999;                          99US-0132487.
PR   07-MAY-1999;                          99US-0132863.
PR   11-MAY-1999;                          99US-0134256.
PR   14-MAY-1999;                          99US-0134218.
PR   14-MAY-1999;                          99US-0134219.
PR   14-MAY-1999;                          99US-0134221.
PR   14-MAY-1999;                          99US-0134370.
PR   18-MAY-1999;                          99US-0134768.
PR   19-MAY-1999;                          99US-0134941.
PR   20-MAY-1999;                          99US-0135124.
PR   21-MAY-1999;                          99US-0135353.
PR   24-MAY-1999;                          99US-0135629.
PR   25-MAY-1999;                          99US-0136021.
PR   27-MAY-1999;                          99US-0136392.
PR   28-MAY-1999;                          99US-0136782.
PR   01-JUN-1999;                          99US-0137222.
PR   03-JUN-1999;                          99US-0137528.
PR   04-JUN-1999;                          99US-0137502.
PR   07-JUN-1999;                          99US-0137724.
PR   08-JUN-1999;                          99US-0138094.
PR   10-JUN-1999;                          99US-0138540.
PR   10-JUN-1999;                          99US-0138847.
PR   14-JUN-1999;                          99US-0139119.
PR   16-JUN-1999;                          99US-0139452.
PR   16-JUN-1999;                          99US-0139453.
PR   17-JUN-1999;                          99US-0139492.
PR   18-JUN-1999;                          99US-0139454.
PR   18-JUN-1999;                          99US-0139455.
PR   18-JUN-1999;                          99US-0139456.
PR   18-JUN-1999;                          99US-0139457.
PR   18-JUN-1999;                          99US-0139458.
PR   18-JUN-1999;                          99US-0139459.
PR   18-JUN-1999;                          99US-0139460.
PR   18-JUN-1999;                          99US-0139461.
PR   18-JUN-1999;                          99US-0139462.
PR   18-JUN-1999;                          99US-0139463.
PR   18-JUN-1999;                          99US-0139750.
PR   18-JUN-1999;                          99US-0139763.
PR   21-JUN-1999;                          99US-0139817.
PR   22-JUN-1999;                          99US-0139899.
PR   23-JUN-1999;                          99US-0140353.
PR   23-JUN-1999;                          99US-0140354.
PR   24-JUN-1999;                          99US-0140695.
PR   28-JUN-1999;                          99US-0140823.
PR   29-JUN-1999;                          99US-0140991.
PR   30-JUN-1999;                          99US-0141287.
PR   01-JUL-1999;                          99US-0141842.
PR   01-JUL-1999;                          99US-0142154.
PR   02-JUL-1999;                          99US-0142055.
PR   06-JUL-1999;                          99US-0142390.
PR   08-JUL-1999;                          99US-0142803.
PR   09-JUL-1999;                          99US-0142920.
PR   12-JUL-1999;                          99US-0142977.
PR   13-JUL-1999;                          99US-0143542.
PR   14-JUL-1999;                          99US-0143524.
PR   15-JUL-1999;                          99US-0144005.
PR   16-JUL-1999;                          99US-0144085.
PR   16-JUL-1999;                          99US-0144086.
PR   19-JUL-1999;                          99US-0144325.
PR   19-JUL-1999;                          99US-0144331.
PR   19-JUL-1999;                          99US-0144332.
PR   19-JUL-1999;                          99US-0144333.
PR   19-JUL-1999;                          99US-0144334.
PR   19-JUL-1999;                          99US-0144335.
PR   20-JUL-1999;                          99US-0144352.
PR   20-JUL-1999;                          99US-0144632.
PR   20-JUL-1999;                          99US-0144884.
PR   21-JUL-1999;                          99US-0144814.
PR   21-JUL-1999;                          99US-0145086.
PR   21-JUL-1999;                          99US-0145088.
PR   22-JUL-1999;                          99US-0145085.
PR   22-JUL-1999;                          99US-0145087.
PR   22-JUL-1999;                          99US-0145089.
PR   22-JUL-1999;                          99US-0145192.
PR   23-JUL-1999;                          99US-0145145.
PR   23-JUL-1999;                          99US-0145218.
PR   23-JUL-1999;                          99US-0145224.
PR   26-JUL-1999;                          99US-0145276.
PR   27-JUL-1999;                          99US-0145913.
```

us-09-540-234-1.rag

```
PR    25-OCT-1999;    99US-0161406.
PR    26-OCT-1999;    99US-0161359.
PR    26-OCT-1999;    99US-0161360.
PR    26-OCT-1999;    99US-0161361.
PR    28-OCT-1999;    99US-0161920.
PR    28-OCT-1999;    99US-0161992.
PR    28-OCT-1999;    99US-0161993.
PR    29-OCT-1999;    99US-0162142.

alignment_scores:
          Quality:   73.00          Length:     128
          Ratio:      0.936         Gaps:         5
Percent Similarity: 60.938   Percent Identity: 24.219

alignment_block:
US-09-540-234-1 x AAG06844   ..

Align seg 1/1   to: AAG06844    from:  1   to: 155

   2 GGTCCACGGCCCTTATGTATGCAGGAGCGCCTAGTGGTCATCTGAAGG   51
     |||  :::|||||||:|::::|: ::::|:  :::||| :::|||||:::
  24 GlySerTyrValLeuPheLeuAlaGlyLysGlyHisThrHisSe   40

  52 AAAGGACAAAGACACCTCCAAGCCGGACACGCCTACTGCTGGCCACGG  101
     :::||| :::|::|||  ||||||||||:::||||||| :::||||||
  40 rHisAsnGlnProMetGluLys..........MetAlaValAlaGlyLeuV   54

 102 CATAATGGAAAACAGTCCCTGCCCCTTTGAGGGTCGTCTCAAACATGCCC  151
     ::  |||||||||||  :::::::|::::||:: :::|:::::::::|||
  54 al.......LeuValProAlaLeuSerProCysAlaThrThrLeuPro   67

 152 CTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAACGCCAATG  201
     |||::|  :::|::||||||::||||:|::||||||||||||||:||||
  68 Val..........PheLeu.AlaValGlyAsnSerLysS   77

 202 GGCTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGTGTAGCGCCTTG  251
     :::::  ||  ::| :::::||||||||||||||||| :::||||||| :::
  77 erMetMetValLeuAlaIleIleValLeuPheSerThrIleLeuVal   93

 252 ATGACTTCCTTGCTGTACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCG  301
     |||||||||||| ||  ::  |||  :::  ||| :::
  94 MetThrSerLeuValAlaLeuSerPheTyrGlyAlaSerGlnLeuLysPh  110

 302 TGCAGCACTCGAAGACTTCCTTGCTGTACTCCGCTGGCACATAGTCTGTT  351
     :::|||  :::|||::::::|||||:::|||||||||||||::|||:|||
 110 eHisTrpValGluArgTyrAspLysLeuValGlySerValLeuCysL  127

 352 TGCTGCCTGGTCTGGGGACACCTGTGTACTAC  383
     ||:::  |||  :::  ||| :::::||:||||
 127 euVal...GlyIleLeuThrLeuLeuPheHis  136

seq_name: /SIDS8/gcgdata/geneseqp/geneseq/AA2000.DAT:AAG06843

seq_documentation_block:
ID   AAG06843 standard; Protein; 225 AA.
XX
AC   AAG06843;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 3767.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
```

Code	Date	Application No.
XX	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	21-JUN-1999;	99US-0139750.
PR	22-JUN-1999;	99US-0139763.
PR	23-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	24-JUN-1999;	99US-0140354.
PR	28-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	02-JUL-1999;	99US-0142154.
PR	06-JUL-1999;	99US-0142055.
PR	08-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.

Code	Date	Application No.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	20-JUL-1999;	99US-0144334.
PR	20-JUL-1999;	99US-0144335.
PR	21-JUL-1999;	99US-0144352.
PR	21-JUL-1999;	99US-0144632.
PR	22-JUL-1999;	99US-0144884.
PR	22-JUL-1999;	99US-0144814.
PR	22-JUL-1999;	99US-0145086.
PR	23-JUL-1999;	99US-0145088.
PR	23-JUL-1999;	99US-0145085.
PR	26-JUL-1999;	99US-0145087.
PR	27-JUL-1999;	99US-0145089.
PR	27-JUL-1999;	99US-0145192.
PR	28-JUL-1999;	99US-0145145.
PR	02-AUG-1999;	99US-0145218.
PR	02-AUG-1999;	99US-0145224.
PR	03-AUG-1999;	99US-0145276.
PR	04-AUG-1999;	99US-0145913.
PR	05-AUG-1999;	99US-0145918.
PR	06-AUG-1999;	99US-0145919.
PR	09-AUG-1999;	99US-0145951.
PR	09-AUG-1999;	99US-0146386.
PR	10-AUG-1999;	99US-0146388.
PR	11-AUG-1999;	99US-0146389.
PR	12-AUG-1999;	99US-0147038.
PR	13-AUG-1999;	99US-0147204.
PR	13-AUG-1999;	99US-0147302.
PR	16-AUG-1999;	99US-0147192.
PR	17-AUG-1999;	99US-0147260.
PR	18-AUG-1999;	99US-0147303.
PR	20-AUG-1999;	99US-0147416.
PR	20-AUG-1999;	99US-0147493.
PR	23-AUG-1999;	99US-0147935.
PR	25-AUG-1999;	99US-0148171.
PR	26-AUG-1999;	99US-0148319.
PR	27-AUG-1999;	99US-0148341.
PR	27-AUG-1999;	99US-0148565.
PR	30-AUG-1999;	99US-0148684.
PR	31-AUG-1999;	99US-0149368.
PR	01-SEP-1999;	99US-0149175.
PR	07-SEP-1999;	99US-0149426.
PR	10-SEP-1999;	99US-0149722.
PR	13-SEP-1999;	99US-0149723.
PR	15-SEP-1999;	99US-0149929.
PR	16-SEP-1999;	99US-0149902.
PR	20-SEP-1999;	99US-0149930.
PR	22-SEP-1999;	99US-0150566.
PR	23-SEP-1999;	99US-0150884.
PR	24-SEP-1999;	99US-0151080.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.

```
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
           Quality:  73.00            Length:      128
             Ratio:   0.936             Gaps:        5
Percent Similarity:  60.938  Percent Identity:   24.219

alignment_block::
US-09-540-234-1 x AAG06843

Align seg 1/1 to: AAG06843 from: 1  to: 225

   2 GGTCCACGGGCCCTTATGTATGCAGGAGCCGCCTAGTGGGTCATCTGAAGG  51
     ::    |::::|:::: ::|||  :::||  |||||||:||||:::|||
  94 GlySerTyrValLeuLeuPheLeuAlaGlyLysGlyHisThrHisSe     110

  52 AAAGGACAAAGACACCTCCAAGCCGACACGGCTACTGAGCTGCCACGG    101
     : :::::::   ::|||            :::||||||||
 110 rHisAsnGlnProMetGluLys.........MetAlaValAlaGlyLeuV  124

 102 CATAATGGAAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAACATGCCC  151
     ::              ||||||||||    :::||::::||||
 124 al.......LeuValProAlaLeuSerProCysAlaThrThrLeuPro    137

 152 CTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAACGCCAATG   201
     :::          |||::   :::|||:|:||:::
 138 Val..............PheLeu.AlaValGlyAsnSerLysS        147

 202 GCGTCGCATGTTATCCGAGCCATTCGTGTGTCGTTTACTGTGTAGCGCTTG 251
     :::::  ||||||||||  ::||||||||||||:::
 147 erMetValLeuAlaIleIleValLeuLeuPheSerThrIleLeuVal    163

 252 ATGACTTCCTTGCTGTACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCG  301
     ||||||||||  ::::    |||    :::|||
 164 MetThrSerLeuValAlaLeuSerPheTyrGlyAlaSerGlnLeuLysPh  180

 302 TGCAGCACTCGAAGACTTCCTTGCTCGTCACTCGGCTCGGCAGCATAGTCTGTT 351
     ::|||    ::    ::|||:|:||||:||||
 180 eHisTrpValGluArgTyrAspLysLysLeuValGlySerValLeuCysL   197
```

```
 352 TGCTGCCTGGTCTGGGACACACCTGTGTACTAC 383
     |:::: |||    :::||||||
 197 euVal...GlyIleLeuThrLeuPheHis 206
```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:AAY01519

```
alignment_scores:
           Quality:  72.50            Length:       96
             Ratio:   1.318             Gaps:        5
Percent Similarity:  57.292  Percent Identity:   28.125

alignment_block:
US-09-540-234-1/rev x AAY01519

Align seg 1/1 to: AAY01519 from: 1  to: 1462

 443 TCGAAACAGACACAGTACATG.........GCTAAGAAGTAGTAC...CA 406
     |||||| ::|||:|||:||    |||||||||| |||   :::
 801 SerAsnValArgGluTyrGlnLysAlaArgAlaLysLysTyrIleGl    817

 405 CTACAAAGGGAAACAGAGTACTGTAGTACACAGGTGTCCCCAGACCAGGC 356
     ::::::::|||:||:|||||    |||   ::::|||
 817 uHisArgSerAsnGluThrSerValIleHisSerLeuProPheGlyGluG  834

 355 AGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTGC 306
     ||   :::|||||||||||  |||     |||||||||
 834 lnThrPheArgProArgGlyMetLeuTyr......GluCysGlnGluCys  848

 305 ......TGCACGGTAAACAGCACTCAAATGGCTCCAAGCGGGAGTACAGCA 262
     ||||||||:||:|||:||||||||  :::
 849 GlyGluCysPheAlaHisSerSerAspLeuThr.................  859

 261 AGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCGATA 212
     ::||||||    :::|||||||
                                                          I
```

```
860  ...GluHisGlnLysIleHisAspArgGluLysProSerGlySerArgA 875

211 ACATCGACGCGCCATTGGCGTTGTTAGTCTTGTACAAACT 174
    ||  :::   |||  |||||||::: |||
875 snTyrGluTrpSerValIleArgSerLeuAlaProThr 887

seq_name: /SIDS8/gcgdata/geneseqp/AA1996.DAT:AAR97246

seq_documentation_block:
ID   AAR97246 standard; Protein; 4472 AA.
XX
AC   AAR97246;
XX
DT   07-JAN-1997  (first entry)
XX
DE   Virulence gene cluster polypeptide product.
XX
KW   Mutant; adaptation; virulence factor; identification; screening;
KW   vaccine; drugs; infection; treatment.
XX
OS   Salmonella typhimurium.
XX
FH   Key             Location/Qualifiers
FT   Region
FT                   /note= "All x's in this sequence correspond to
FT                   termination codons in the virulence gene
FT                   cluster sequence given in AAT09224."
XX
PN   WO9617951-A2.
XX
PD   13-JUN-1996.
XX
PF   11-DEC-1995; 95WO-GB02875.
XX
PR   05-MAY-1995; 95GB-0009239.
PR   09-DEC-1994; 94GB-0024921.
PR   31-JAN-1995; 95GB-0001881.
XX
PA   (RPMS-) RPMS TECHNOLOGY LTD.
XX
PI   Holden DW;
XX
DR   WPI; 1996-287194/29.
XX
PT   Identifying virulence genes in microorganisms - by introducing
PT   mutants with insertion inactivated genes into environment and
PT   retrieval and analysis of mutants
XX
PS   Claim 51; Figure 11; 131pp; English.
XX
CC   A method for identifying a microorganism having a reduced adaptation
CC   to a particular environment comprising the steps of: (1) providing a
CC   plurality of microorganisms each of which is independently mutated by
CC   the insertional inactivation of a gene with a nucleic acid comprising
CC   a unique marker sequence so that each mutant contains a different
CC   marker sequence, or clones of the said microorganism; (2) providing
CC   individually a stored sample of each mutant produced by step (1) and
CC   providing individually stored nucleic acid comprising the unique
CC   marker sequence from each individual mutant; (3) introducing a
CC   plurality of mutants produced by step (1) into the said particular
CC   environment and allowing those microorganisms which are able to do so
CC   to grow in the said environment; (4) retrieving microorganisms from
CC   the said environment or a selected part thereof and isolating the
CC   nucleic acid from the retrieved microorganism; (5) comparing any
CC   marker sequences in the nucleic acid isolated in step (4) to the
CC   unique marker sequence of each individual mutant stored as in step
CC   (2); and (6) selecting an individual mutant which does not contain any
CC   of the marker sequences as isolated in step (4).  The products and
CC   methods can be used for identifying virulence genes in microorganisms.
CC   The mutant microorganisms can be used in vaccines or to screen for
CC   drugs which reduce virulence or compounds useful for preventing,
CC   ameliorating or treating infections in animals or plants.
XX
```

```
SQ   Sequence   4472 AA;

alignment_scores:
        Quality:   72.50                 Length:    136
          Ratio:    1.229                  Gaps:       9
     Percent Similarity:  43.382      Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR97246     .

Align seq 1/1 to: AAR97246  from: 1  to:4472

  61 AGACACCTCCAAGCCGACACGGCTACTGTAGCTGGCACGGCATAATGGA 110
     |||::         ::|||::    ::|||::  ::||||||||::::|||
2023 ArgArgAlaThrSerGlyTyrPhePheCysSerTrpArgGly........ 2036

 111 AAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACG 160
                                          ||||||||||
2037 ..........................GluLysHisAlaSer.......... 2041

 161 TCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAA.......... 199
     |||::||||||  |||::|||      ||:::|||
2042 ..GlyAspAlaVal***CysAlaArgArg.ArgGlnCysSerGlyVa 2057

 200   TGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGCTGTTT 236
       ||| |||   |||  ||  |||| ||||    :|||::|||
2057 lAsnTrp***ThrTrpThrArgSerProArgIleHisArg......... 2070

 237 ACTGTGTAGCGCTTGATGACTTCCTTGCGTGCTACTCCGCTTGGAGCCA. 283
     |||  :::||||||      |||  ||   :::|||
2071                  PheTyrThrVal***Arg...AspProLys 2079

 284 .......TTTGAGTGCTGTTACCGTGCGAGCACTCGAAGACTT........ 319
          :::||||||        :::        ||||||
2080 ThrLeuCysHisCysCysArgAsnLeu***GlnThrArgLeuArgAlaAr 2096

 320 .......CCTTGCTGTACTCCGCTGGCAGCATAGTCTGTTTGCT...GCCT 359
           |||           |||   ::::|||      :::|||:::|||
2096 gGluGlyAlaValCysGlyHisHisAspSerArgIlePheSerArg***T 2113

 360 GGTCTGGGGACACCTGTGTACTACACAGTACTTCGTTTCCCTTGTAGTGGT 409
     ||:::::::||:::  :::     :::         ::||||||:::|||
2113 rpLysAlaSerArgLeuAlaCysArgLeuThrAspAlaLeuCysGlnGly 2129

 410 ACTACT 415
     ||||||
2130 ArgThr 2131
```

```
seq_name: /SIDS8/gcgdata/geneseqp/AA2000.DAT:AAG17227

seq_documentation_block:
ID   AAG17227 standard; Protein; 307 AA.
XX
AC   AAG17227;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 18167.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
```

```
PR 05-MAR-1999; 99US-0123180.          PR 14-JUL-1999; 99US-0143624.
PR 09-MAR-1999; 99US-0123548.          PR 15-JUL-1999; 99US-0144005.
PR 23-MAR-1999; 99US-0125788.          PR 16-JUL-1999; 99US-0144085.
PR 25-MAR-1999; 99US-0126264.          PR 16-JUL-1999; 99US-0144086.
PR 29-MAR-1999; 99US-0126785.          PR 19-JUL-1999; 99US-0144325.
PR 01-APR-1999; 99US-0127462.          PR 19-JUL-1999; 99US-0144331.
PR 06-APR-1999; 99US-0128234.          PR 19-JUL-1999; 99US-0144332.
PR 08-APR-1999; 99US-0128714.          PR 19-JUL-1999; 99US-0144333.
PR 16-APR-1999; 99US-0129845.          PR 19-JUL-1999; 99US-0144334.
PR 19-APR-1999; 99US-0130077.          PR 19-JUL-1999; 99US-0144335.
PR 21-APR-1999; 99US-0130449.          PR 20-JUL-1999; 99US-0144352.
PR 23-APR-1999; 99US-0130510.          PR 20-JUL-1999; 99US-0144632.
PR 23-APR-1999; 99US-0130891.          PR 20-JUL-1999; 99US-0144884.
PR 28-APR-1999; 99US-0131149.          PR 21-JUL-1999; 99US-0144814.
PR 30-APR-1999; 99US-0132048.          PR 21-JUL-1999; 99US-0145086.
PR 30-APR-1999; 99US-0132407.          PR 21-JUL-1999; 99US-0145088.
PR 04-MAY-1999; 99US-0132484.          PR 22-JUL-1999; 99US-0145085.
PR 05-MAY-1999; 99US-0132485.          PR 22-JUL-1999; 99US-0145087.
PR 06-MAY-1999; 99US-0132486.          PR 22-JUL-1999; 99US-0145089.
PR 06-MAY-1999; 99US-0132487.          PR 22-JUL-1999; 99US-0145192.
PR 07-MAY-1999; 99US-0132663.          PR 23-JUL-1999; 99US-0145145.
PR 11-MAY-1999; 99US-0134256.          PR 23-JUL-1999; 99US-0145218.
PR 14-MAY-1999; 99US-0134218.          PR 23-JUL-1999; 99US-0145224.
PR 14-MAY-1999; 99US-0134219.          PR 26-JUL-1999; 99US-0145276.
PR 14-MAY-1999; 99US-0134221.          PR 27-JUL-1999; 99US-0145913.
PR 18-MAY-1999; 99US-0134370.          PR 27-JUL-1999; 99US-0145918.
PR 18-MAY-1999; 99US-0134768.          PR 27-JUL-1999; 99US-0145919.
PR 19-MAY-1999; 99US-0134941.          PR 28-JUL-1999; 99US-0145951.
PR 20-MAY-1999; 99US-0135124.          PR 02-AUG-1999; 99US-0146386.
PR 21-MAY-1999; 99US-0135353.          PR 02-AUG-1999; 99US-0146388.
PR 24-MAY-1999; 99US-0135629.          PR 03-AUG-1999; 99US-0146389.
PR 25-MAY-1999; 99US-0136021.          PR 04-AUG-1999; 99US-0147038.
PR 27-MAY-1999; 99US-0136392.          PR 04-AUG-1999; 99US-0147204.
PR 28-MAY-1999; 99US-0136782.          PR 05-AUG-1999; 99US-0147302.
PR 01-JUN-1999; 99US-0137222.          PR 06-AUG-1999; 99US-0147192.
PR 03-JUN-1999; 99US-0137528.          PR 06-AUG-1999; 99US-0147260.
PR 04-JUN-1999; 99US-0137502.          PR 09-AUG-1999; 99US-0147303.
PR 07-JUN-1999; 99US-0137724.          PR 09-AUG-1999; 99US-0147416.
PR 08-JUN-1999; 99US-0138094.          PR 10-AUG-1999; 99US-0147493.
PR 10-JUN-1999; 99US-0138540.          PR 11-AUG-1999; 99US-0147935.
PR 10-JUN-1999; 99US-0138847.          PR 12-AUG-1999; 99US-0148171.
PR 14-JUN-1999; 99US-0139119.          PR 13-AUG-1999; 99US-0148319.
PR 16-JUN-1999; 99US-0139452.          PR 16-AUG-1999; 99US-0148341.
PR 16-JUN-1999; 99US-0139453.          PR 17-AUG-1999; 99US-0148565.
PR 17-JUN-1999; 99US-0139492.          PR 18-AUG-1999; 99US-0148684.
PR 18-JUN-1999; 99US-0139454.          PR 20-AUG-1999; 99US-0149368.
PR 18-JUN-1999; 99US-0139455.          PR 20-AUG-1999; 99US-0149175.
PR 18-JUN-1999; 99US-0139456.          PR 23-AUG-1999; 99US-0149426.
PR 18-JUN-1999; 99US-0139457.          PR 23-AUG-1999; 99US-0149723.
PR 18-JUN-1999; 99US-0139458.          PR 25-AUG-1999; 99US-0149929.
PR 18-JUN-1999; 99US-0139459.          PR 26-AUG-1999; 99US-0149902.
PR 18-JUN-1999; 99US-0139460.          PR 27-AUG-1999; 99US-0149930.
PR 18-JUN-1999; 99US-0139461.          PR 27-AUG-1999; 99US-0150566.
PR 18-JUN-1999; 99US-0139462.          PR 27-AUG-1999; 99US-0150884.
PR 18-JUN-1999; 99US-0139463.          PR 30-AUG-1999; 99US-0151065.
PR 18-JUN-1999; 99US-0139750.          PR 31-AUG-1999; 99US-0151066.
PR 21-JUN-1999; 99US-0139763.          PR 01-SEP-1999; 99US-0151080.
PR 22-JUN-1999; 99US-0139817.          PR 07-SEP-1999; 99US-0151303.
PR 23-JUN-1999; 99US-0139899.          PR 10-SEP-1999; 99US-0151438.
PR 23-JUN-1999; 99US-0140353.          PR 13-SEP-1999; 99US-0151930.
PR 24-JUN-1999; 99US-0140354.          PR 15-SEP-1999; 99US-0152363.
PR 28-JUN-1999; 99US-0140695.          PR 16-SEP-1999; 99US-0153070.
PR 29-JUN-1999; 99US-0140823.          PR 20-SEP-1999; 99US-0153758.
PR 30-JUN-1999; 99US-0140991.          PR 22-SEP-1999; 99US-0154018.
PR 01-JUL-1999; 99US-0141287.          PR 23-SEP-1999; 99US-0154039.
PR 01-JUL-1999; 99US-0141842.          PR 24-SEP-1999; 99US-0154779.
PR 02-JUL-1999; 99US-0142154.          PR 28-SEP-1999; 99US-0155139.
PR 06-JUL-1999; 99US-0142055.          PR 29-SEP-1999; 99US-0155486.
PR 08-JUL-1999; 99US-0142390.          PR 04-OCT-1999; 99US-0155659.
PR 09-JUL-1999; 99US-0142803.          PR          99US-0156458.
PR 12-JUL-1999; 99US-0142920.          PR          99US-0156596.
PR 12-JUL-1999; 99US-0142977.          PR          99US-0157117.
PR 13-JUL-1999; 99US-0143542.
```

```
                                                        |||   |||:::::::          :::|||||||||
                                            252 GlnLeuGlyGluValIleTyrValLysIleProAlaThr 264

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAG17226

seq_documentation_block:
ID   AAG17226 standard; Protein; 308 AA.
XX
AC   AAG17226;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 18166.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   23-APR-1999;  99US-0130891.
PR   28-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   18-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.

PR   05-OCT-1999;  99US-0157753.
PR   06-OCT-1999;  99US-0157865.
PR   07-OCT-1999;  99US-0158029.
PR   08-OCT-1999;  99US-0158232.
PR   12-OCT-1999;  99US-0158369.
PR   13-OCT-1999;  99US-0159293.
PR   13-OCT-1999;  99US-0159294.
PR   13-OCT-1999;  99US-0159295.
PR   14-OCT-1999;  99US-0159329.
PR   14-OCT-1999;  99US-0159330.
PR   14-OCT-1999;  99US-0159331.
PR   14-OCT-1999;  99US-0159637.
PR   14-OCT-1999;  99US-0159638.
PR   18-OCT-1999;  99US-0159584.
PR   21-OCT-1999;  99US-0160741.
PR   21-OCT-1999;  99US-0160767.
PR   21-OCT-1999;  99US-0160768.
PR   21-OCT-1999;  99US-0160770.
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   26-OCT-1999;  99US-0161406.
PR   26-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.

alignment_scores:
            Quality:  71.50          Length:     131
            Ratio:    0.979          Gaps:       7
Percent Similarity:  55.725   Percent Identity:  24.427

alignment_block:
US-09-540-234-1/rev x AAG17227  :

Align seq 1/1  to: AAG17227  from: 1 to: 307

407 CACTACAAAGGGAAACGAAGGAGTACTGTAGTACACAGAGTGTCCCCAGACCAG 358
    ||||||::::|  |||:||:::  |||   |||||:: |:|     ::::|||
136 HisTyrSerValArgGlyAlaGlyAlaLysValValThrAspProSerThrGl 152

357 GCAGCAAACAGACTATGCT................GCCAGCGAGTACA 326
    :::::|    |||:::      |||||||  :
152 yArgSerLysGlyTyrGlyPheValLysPheAlaGluSerGluArgA 169

325 GCAAGGAAGTCTTCGAGTGC..........TGCACGGTAAACAGCACT 288
    :::: :::  |||   |||               |||:::::::::
169 snArgAlaMetAlaGluMetAsnGlyLeuTyrCysSerThrArgProMet 185

287 CAAATGGCTCCAAGC......................GGAGTACAGCAAGGAAG 256
    :::::::::   :::                       ||||||||||   :::
186 ArgIleSerAlaAlaThrProLysLysAsnValGlyValGlnGln.GlnT 202

255 TCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCGATA....... 212
    :::  :::   ||||| |||||   |||||
202 yrValThrLysValThrValProSerAlaValAlaAlaProValGlnAla 218

211 ..ACATCGACGCCATTGGCGTTGCTGTTTGTTAGTTCTTGTACAAACTACGGCTTCA 165
    ::::::::    |||  :::  ::||||||:: ::||||||::
219 TyrValAlaProProGluSerAspVal.ThrCysThrThrIleSerValA 235

164 CCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGGCA...... 121
    :::::::::::::      |||         ||||||::|||:::|||
235 laAsnLeuAspGlnAsnValThrGluGluGluLysLysValAlaPheSer 251

120 ......GGGACTGTTTCCATTATGCCGTGCCAGCTACA 88
```

```
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  20-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  21-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.
PR  22-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  05-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  06-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  09-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  10-AUG-1999;  99US-0147935.
PR  11-AUG-1999;  99US-0148171.
PR  12-AUG-1999;  99US-0148319.
PR  13-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  16-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.

PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  22-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  25-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  26-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  28-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  29-OCT-1999;  99US-0161993.
PR             ;  99US-0162142.


alignment_scores:
                    Quality:  71.50         Length:  131
                    Ratio:    0.979         Gaps:    7
Percent Similarity:  55.725   Percent Identity:  24.427

alignment_block:
US-09-540-234-1/rev x AAG17226  ..

Align seg 1/1  to: AAG17226  from: 1  to: 308

407 CACTACAAAGGGAAACGAAGTACTGTAGTAGTACACAGGTGTCCCCAGACCAG 358
    ||||||:||:::::     |||:::::::      |||
```

```
137 HisTyrSerSerValArgGlyAlaLysValValThrAspProSerThrGl 153
        :::::::    :::                          ::::
357 GCAGCAAACAGCTATGCT................GCCACCGGAGTACA 326
        ::::::      ||||||                  ||||||
153 yArgSerLysGlyTyrGlyPheValLysPheAlaGluGluSerGluArgA 170
        :: :::                                :
325 GCAAGGAAGTCTTCGAGTGC...........TGCACCGGTAAACAGCACT 288
        :: ::  :::  |||                  |||:::::::::
170 snArgAlaMetAlaGluMetAsnGlyLeuTyrCysSerThrArgProMet 186

287 CAAATGGCTCCAAGC................GGAGTACAGCAAGGAAG 256
        ::::::::::  :::                  ||||||||||| :::
187 ArgIleSerAlaAlaThrProLysLysAsnValGlyValGlnGln.GlnT 203

255 TCATCAAGCGCTACACAGTAAACAGCACACGCGAATGGCTCCGATA...... 212
        :: ::::||||||   |||:::||||||||   ||||||
203 yrValThrLysValProSerAlaValAlaValAlaProValGlnAla 219

211 ...ACATCGACGCGCCATTGCGGTTGTTAGTCTTGTACAAACTACGGCTTCA 165
        :: ::::  |||  ::: ::::|||||:::  :: :::
220 TyrValAlaProGluProGluSerAspVal.ThrCysThrThrIleSerValA 236

164 CCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGGCA...... 121
        :::::::::::::::: ::|||   ||||||||:::|||
236 laAsnLeuAspGlnAsnValThrGluGluGluLeuLeuLysValAlaPheSer 252

120 .....GGGACTGTTTTCCATTATGCCGTGCCAGCTACA 88
        |||  ||:::::::::         ::|||||||||
253 GlnLeuGlyGluValIleTyrValLysIleProAlaThr 265
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG17225

seq_documentation_block:
ID AAG17225 standard; Protein; 310 AA.
XX
AC AAG17225;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 18165.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999;   99US-0121825.
PR 05-MAR-1999;   99US-0123180.
PR 09-MAR-1999;   99US-0123548.
PR 23-MAR-1999;   99US-0125788.
PR 25-MAR-1999;   99US-0126264.
PR 29-MAR-1999;   99US-0126785.
PR 01-APR-1999;   99US-0127462.
PR 06-APR-1999;   99US-0128234.
PR 08-APR-1999;   99US-0128714.
PR 16-APR-1999;   99US-0129845.
PR 19-APR-1999;   99US-0130077.
PR 21-APR-1999;   99US-0130449.
PR 23-APR-1999;   99US-0130510.
PR 23-APR-1999;   99US-0130891.
PR 28-APR-1999;   99US-0131449.
PR 30-APR-1999;   99US-0132048.
PR 30-APR-1999;   99US-0132407.
PR 04-MAY-1999;   99US-0132484.
PR 05-MAY-1999;   99US-0132485.
PR 06-MAY-1999;   99US-0132486;
PR 06-MAY-1999;   99US-0132487.
PR 07-MAY-1999;   99US-0132863.
PR 11-MAY-1999;   99US-0134256.
PR 14-MAY-1999;   99US-0134218.
PR 14-MAY-1999;   99US-0134219.
PR 14-MAY-1999;   99US-0134221.
PR 18-MAY-1999;   99US-0134370.
PR 19-MAY-1999;   99US-0134768.
PR 19-MAY-1999;   99US-0134941.
PR 20-MAY-1999;   99US-0135124.
PR 21-MAY-1999;   99US-0135353.
PR 24-MAY-1999;   99US-0135629.
PR 25-MAY-1999;   99US-0136021.
PR 27-MAY-1999;   99US-0136392.
PR 28-MAY-1999;   99US-0136782.
PR 01-JUN-1999;   99US-0137222.
PR 03-JUN-1999;   99US-0137528.
PR 04-JUN-1999;   99US-0137502.
PR 07-JUN-1999;   99US-0137724.
PR 08-JUN-1999;   99US-0138094.
PR 10-JUN-1999;   99US-0138540.
PR 10-JUN-1999;   99US-0138847.
PR 14-JUN-1999;   99US-0139119.
PR 16-JUN-1999;   99US-0139452.
PR 16-JUN-1999;   99US-0139453.
PR 17-JUN-1999;   99US-0139492.
PR 18-JUN-1999;   99US-0139454.
PR 18-JUN-1999;   99US-0139455.
PR 18-JUN-1999;   99US-0139456.
PR 18-JUN-1999;   99US-0139458.
PR 18-JUN-1999;   99US-0139459.
PR 18-JUN-1999;   99US-0139460.
PR 18-JUN-1999;   99US-0139461.
PR 18-JUN-1999;   99US-0139462.
PR 18-JUN-1999;   99US-0139463.
PR 18-JUN-1999;   99US-0139750.
PR 18-JUN-1999;   99US-0139763.
PR 21-JUN-1999;   99US-0139817.
PR 22-JUN-1999;   99US-0139899.
PR 23-JUN-1999;   99US-0140353.
PR 23-JUN-1999;   99US-0140354.
PR 24-JUN-1999;   99US-0140695.
PR 28-JUN-1999;   99US-0140823.
PR 29-JUN-1999;   99US-0140991.
PR 30-JUN-1999;   99US-0141287.
PR 01-JUL-1999;   99US-0141842.
PR 01-JUL-1999;   99US-0142154.
PR 02-JUL-1999;   99US-0142055.
PR 06-JUL-1999;   99US-0142390.
PR 08-JUL-1999;   99US-0142803.
PR 09-JUL-1999;   99US-0142920.
PR 12-JUL-1999;   99US-0142977.
PR 13-JUL-1999;   99US-0143542.
PR 14-JUL-1999;   99US-0143624.
PR 15-JUL-1999;   99US-0144005.
PR 16-JUL-1999;   99US-0144085.
PR 16-JUL-1999;   99US-0144086.
PR 19-JUL-1999;   99US-0144325.
PR 19-JUL-1999;   99US-0144331.
PR 19-JUL-1999;   99US-0144332.
PR 19-JUL-1999;   99US-0144333.
PR 19-JUL-1999;   99US-0144334.
PR 19-JUL-1999;   99US-0144335.
PR 20-JUL-1999;   99US-0144352.
PR 20-JUL-1999;   99US-0144632.
PR 20-JUL-1999;   99US-0144884.
PR 21-JUL-1999;   99US-0144814.
PR 21-JUL-1999;   99US-0145086.
PR 21-JUL-1999;   99US-0145088.
PR 22-JUL-1999;   99US-0145085.
PR 22-JUL-1999;   99US-0145087.
PR 22-JUL-1999;   99US-0145089;

PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   10-AUG-1999;   99US-0147935.
PR   11-AUG-1999;   99US-0148171.
PR   12-AUG-1999;   99US-0148319.
PR   13-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151103.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.

PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

alignment_scores:
           Quality: 71.50   Length: 131
           Ratio: 0.979   Gaps: 7
Percent Similarity: 55.725   Percent Identity: 55.725

alignment_block:
US-09-540-234-1/rev x AAG17225

Align seg 1/1 to: AAG17225 from: 1 to: 310

407 CACTACAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAG 358
        |||||:::||| |||:::::: ||| ||||||||:::|||
139 HisTyrSerSerValArgGlyAlaLysValValThrAspProSerThrG1 155

357 GCAGCAAACAGACTATGCT..................GCCAGCGAGTACA 326
        ::::::: ||||| |||||||
155 yArgSerLysGlyTyrGlyPheValLysPheAlaGluGluSerGluArgA 172

325 GCAAGGAGTCTTCGAGTGC...........TGCACCGTAAACAGCACT 288
        :::::: ||| |||:::::::::
172 snArgAlaMetAlaGluMetAsnGlyLeuTyrCysSerThrArgProMet 188

287 CAAATGGCTCCAAGC......................GGAGTACAGCAAGGAAG 256
        |||||||||||||| |||||||||||||
189 ArgIleSerAlaAlaThrProLysAsnValGlyValGlnGln.GlnT 205

255 TCATCAAGGCGCTACACAGTAAACAGCACACGAATGGCTCCGATA...... 212
        ::: ||||| |||::: |||||||
205 yrValThrLysValProSerAlaValAlaAlaProValGlnAla 221

211 ...ACATCGACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCA 165
       :::::: ||| ::: ::|||||:::::||||||| :::
222 TyrValAlaProProGluSerAspVal.ThrCysThrThrIleSerValA 238

164 CCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGCA...... 121
        ||| ||| || |||||:::||||
238 laAsnLeuAspGlnAsnValThrGluGluGluLysLysLysAlaPheSer 254

120 ......GGGACTGTTTTCCATTATGCCGTGCCAGCTACA 88
        ||| ::|||||||||||
255 GlnLeuGlyGluValIleTyrValLysIleProAlaThr 267

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA1993.DAT:AAR38470

```
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Region          1021..1221
FT                   /note= "Region of homology to Drosophila trithorax".
FT   Region          1462..1570
FT                   /note= "Region of homology to Drosophila trithorax"
FT   Region          3348..3562
FT                   /note= "Region of homology to Drosophila trithorax"
XX
PN   WO9312136-A.
XX
PD   24-JUN-1993.
XX
PF   09-DEC-1992;  92WO-US10930.
XX
PR   11-DEC-1991;  91US-0805093.
PR   27-MAY-1992;  92US-0888839.
PR   30-OCT-1992;  92US-0971094.
XX
PA   (UVJE-) UNIV JEFFERSON THOMAS.
XX
PI   Canaani E,  Croce CM;
XX
DR   WPI; 1993-214090/26.
DR   N-PSDB: AAQ43526.
XX
PT   Detection and treatment of acute leukaemia(s) - using prods.
PT   derived from oligo:nucleotide sequences within the ALL-1 gene of
PT   chromosome 11
XX
PS   Disclosure; Page 29-50; 90pp; English.
XX
CC   This sequence is encoded by the acute lymphoblastic leukemia (ALL-1)
CC   gene of chromosome 11.  The ALL-1 gene was isolated by translocation
CC   breakpoint mapping.  Fragments of the ALL-1 cDNA may be used to
CC   identify chromosomal abnormalities within the ALL-1 gene.  These
CC   fragments may be used in the treatment and diagnosis of human
CC   leukemias such as acute lymphocytic, myelomonocytic, monocytic and
CC   myelogenous leukemia.  ALL-1 protein shows three regions of homology
CC   to the Drosophila trithorax protein.  These regions show 64%, 66% and
CC   82% similarity respectively, to the Drosophila gene.  The third region
CC   of homology constitutes the extreme C-terminus of the two proteins,
CC   both proteins end in an identical sequence.  The first homology region
CC   is cysteine-rich and contains sequence motifs analogous to four zinc
CC   finger domains (3-6) within the trithorax gene.  The second region of
CC   homology is also cysteine-rich and corresponds to zinc fingers 7 and 8
CC   of the Drosophila gene.  The multiple conserved cysteines and
CC   histidines at the 3' end of the motifs allow two or three arrangements
CC   of the putative fingers.  The structure of these cysteine-rich domains
CC   appears to be unique to the trithorax and ALL-1 genes.
XX
SQ   Sequence  3910 AA;
```

alignment_scores:
       Quality: 71.50   Length: 118
       Ratio: 1.192   Gaps: 6
Percent Similarity: 50.847   Percent Identity: 29.661

alignment_block:
US-09-540-234-1/rev x AAR38470 ..

Align seg 1/1 to: AAR38470 from: 1 to: 3910

```
367 CCCAGACCAGGCAGCCAAACAGACTATGCTGCCCAGCGAGTACAGCAGGGAA 318
    |||::: ||||::::|||||  |||||    ||| :||  ::::
3348 ProGlnLeuGlyThrSerGlnThrProSerThrAlaAlaIleThrAlaAl 3364

317 GTCTTCGAGTGCTGCACGGTAAACAGCACTCAA.................. 285
    ||||||    |||||:::::   ||||||||||
3364 aSerSer..IleCysValProSerThrGlnThrThrGlyIleThrArAl 3380
```

```
284 .ATGGCTCCAAGCGGA.....................................GTAC 266
     :::|||||||||                                         |||
3380 aAlaSerProSerGlyGluAlaAspGluHisTyrGlnLeuGlnHisValA 3397

265 AGCAA.......GGAAGTCATCAAGCGGCTACACAGTAAAACAGCAC...... 228
     :::|||       :::||| ::::::||||||:::||||:::||||:::
3397 snGlnLeuAlaSerLysThrGlyIleHisSerSerGlnArgAspLeu 3413

227 ....................ACGAATGGCTCCGATAACATCGACGC 202
                         :::|||   ::::::::::::!|||
3414 AspSerAlaSerGlyProGlnValSerAsnPheThrGlnThrValAspAl 3430

201 CATTGGCGTGTTGTTAGTCTTGTACAA......................ACTA 173
    |       ||::::::||| |||                          :::
3430 aProAsnSerMetGlyLeuGluGlnAsnLysAlaLeuSerSerAlaValG 3447

172 CGGCTTCACCGACGCTCCGAGGGCATGTTTGAGACGACCCTCAAAAGGG 123
    |||||||||||||||     ||||||
3447 lnAlaSerProThrSerProGlyGlySerProSerSerGly 3463

122 CAG 120
    |||
3464 Gln 3464
```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:AAR66462

```
seq_documentation_block:
ID   AAR66462 standard; Protein: 3910 AA.
XX
AC   AAR66462;
XX
DT   22-AUG-1995   (first entry)
XX
DE   ALL-1 (acute lymphocytic leukaemia-1) protein.
XX
KW   Acute lymphoblastic leukaemia; acute nonlymphoblastic leukaemia;
KW   chromosomal translocation; abnormality; detection; t(4:11);
KW   t(9:11); t(11:17); ALL-1; AF-4; AF-9; AF-17; chimeric gene; probe.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Protein         348..3910
FT                   /note= "in the disclosure, numbering of amino acid
FT                   residues starts from Met348 as position 1"
FT   Region          1369..1559
FT                   /note= "cysteine-rich region of homology to
FT                   Drosophila trithorax, contg. sequence
FT                   motifs analogous to zinc finger domains"
FT   Domain          1372..1397
FT                   /label= zinc_finger
FT                   /note= "homologous to zinc finger domain 3 of
FT                   Drosophila trithorax"
FT   Domain          1417..1447
FT                   /label= zinc_finger
FT                   /note= "homologous to zinc finger domain 4 of
FT                   Drosophila trithorax"
FT   Domain          1465..1497
FT                   /label= zinc_finger
FT                   /note= "homologous to zinc finger domain 5 of
FT                   Drosophila trithorax"
FT   Domain          1507..1537
FT                   /label= zinc_finger
FT                   /note= "homologous to zinc finger domain 6 of
FT                   Drosophila trithorax"
FT   Region          1810..1918
FT                   /note= "cysteine-rich region of homology to
FT                   Drosophila trithorax, contg. sequence
FT                   motifs analogous to zinc fingers"
FT   Domain          1814..1844
FT                   /label= zinc_finger
FT                   /note= "homologous to zinc finger domain 7 of
```

```
FT   Domain            1874..1899
FT                     /label= zinc_finger
FT                     /note= "homologous to zinc finger domain 8 of
FT                     Drosophila trithorax"
FT   Region            3696..3910
FT                     /note= "region of homology to Drosophila trithorax
FT                     located at extreme C-terminus of both
FT                     proteins"
XX
PN   WO9426930-A.
XX
PD   24-NOV-1994.
XX
PF   22-APR-1994;    94WO-US04496.
XX
PR   14-MAY-1993;    93US-0062443.
XX
PA   (UYJE-) UNIV JEFFERSON THOMAS.
XX
PI   Canaani E, Croce C;
XX
DR   WPI; 1995-006818/01.
DR   N-PSDB; AAQ75181.
XX
PT   New acute lymphocytic leukaemia gene prods. - used for the
PT   diagnosis and treatment of leukaemias, partic. acute
PT   lymphoblastic or nonlymphoblastic leukaemia
XX
PS   Claim 65; Page 62-79; 207pp; English.
XX
CC   The ALL-1 cDNA was obtained from a series of overlapping clones
CC   spanning 14.7 kb, isolated by screening a human fibroblast library
CC   and a K562 library. The sequence revealed a single, long ORF
CC   predicting a protein of approx. 4000 amino acids. The predicted
CC   amino acid sequence includes regions of homology with the
CC   Drosophila trithorax gene which contain zinc-finger domains. The
CC   multiple conserved cysteines and histidines at the 3' end of the
CC   motifs allow two or three arrangements of the putative fingers.
CC   The ALL-1 gene on chromosome 11 is involved in a series of
CC   chromosomal translocations which are associated with certain
CC   leukaemias.
XX
SQ   Sequence   3910 AA;

alignment_scores:
       Quality:    71.50                 Length:      118
       Ratio:       1.192                 Gaps:          6
Percent Similarity:  50.847     Percent Identity:   29.661

alignment_block:
US-09-540-234-1/rev x AAR66462        ..

Align seg 1/1   to: AAR66462   from: 1  to: 3910

367  CCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCAGCGAGTACAGCAAGGAA  318
     |||::   ||:::::|||||               |||:::        ::
3348 ProGlnLeuGlyThrSerGlnThrProSerThrAlaAlaIleThrAlaAl  3364

317  GTCTTCGAGTGCTGCACGGTAAACAGCACTCAA...............  285
     ||||| |::::::|||  |||::||||
3364 aSerSer..IleCysValLeuProSerThrGlnThrThrGlnGlyIleThrAl  3380

284  .ATGGCTCCAAGCGGA.......................GTAC  266
     |:||||||||||               |||
3380 aAlaSerProSerGlyGluAlaAspGluHisTyrGlnLeuGlnHisValA  3397

265  AGCAA.....GGAAGTCATCAAGCCCTACACAGTAAACAGCAC......  228
     ::|||     :::||||||||||||||:::
3397 snGlnLeuLeuAlaSerLysThrGlyIleHisSerSerGlnArgAspLeu  3413

227  ........................ACGAATGGCTCCGATAACATCGACGC  202
```

```
3414 AspSerAlaSerGlyProGlnValSerAsnPheThrGlnThrValAspAl  3430
     :::|||                        :::::::::::::::|||||
201  CATTGGCCGTGTTAGTCTTGTACAA................ACTA  173
     |  ::   :|:::|||  |||                      :::
3430 aProAsnSerMetGlyLeuGluGluGlnAsnLysAlaLeuSerSerAlaValG  3447
     ||||||||||||||||||||||||||||||||||||||||:::
172  CGGCTTCACCGACGTCTCGAGGGCATGTTTGAGACGACCCTCAAAAGGG  123
     ||||||||||||||||||||||||||||||||::::
3447 lnAlaSerProThrSerProGlyGlySerProSerProSerSerGly  3463
     |||||:::||||
122  CAG  120
     |||
3464 Gln  3464

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:AAR52971

seq_documentation_block:
ID   AAR52971 standard; Protein; 3969 AA.
XX
AC   AAR52971;
XX
DT   27-SEP-1994  (first entry)
XX
DE   Product of the cDNA encoding htrx.
XX
KW   Human; trithorax gene; L01986; diagnosis; treatment;
KW   immunodeficiency; developmental abnormalities; inherited diseases;
KW   cancer; acute lymphocytic leukaemia; myelomonocytic leukaemia.
XX
OS   Homo sapiens.
XX
PN   WO9407502-A.
XX
PD   14-APR-1994.
XX
PF   24-SEP-1993;    93WO-US09087.
XX
PR   30-SEP-1992;    92US-0954112.
PR   13-MAY-1993;    93US-0061376.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Djabali M, Evans GA, Parry P, Selleri L;
XX
DR   WPI; 1994-135206/16.
XX
PT   Nucleic acid encoding a human tri:thorax protein - used to
PT   develop agents for diagnosis and treatment of diseases associated
PT   with disruption of chromosome II at q23
XX
PS   Disclosure; Page 43-54; 68pp; English.
XX
CC   In the course of the construction of a physical map of human
CC   chromosome region 11q23, a region contg. the t(4;11) translocation
CC   breakpoint was cloned. The cloned DNA encoded a protein homologous
CC   to the trithorax gene prod. of Drosophila. The gene may be used for
CC   the diagnosis and treatment of immunodeficiency states,
CC   developmental abnormalities, inherited diseases or cancers, e.g.
CC   acute lymphocytic leukaemia or acute myelomonocytic leukaemia.
CC   See also AAR52972-7.
XX
SQ   Sequence   3969 AA;

alignment_scores:
       Quality:    71.50                 Length:      118
       Ratio:       1.192                 Gaps:          6
Percent Similarity:  50.847     Percent Identity:   29.661

alignment_block:
US-09-540-234-1/rev x AAR52971        ..
```

Align seg 1/1 to: AAR52971 from: 1 to: 3969

```
  367 CCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAA 318
      |||:::||||||||||    :::::|||||||    |||:::
 3406 ProGlnLeuGlyThrSerGlnThrProSerThrAlaAlaIleThrAlaAl 3422

  317 GTCTTCGAGTGCTGCACGGTAAACAGCACTCAA................. 285
      ||||||||  |||:::::: |||||||||
 3422 aSerSer..IleCysValLeuProSerThrGlnThrThrGlyIleThrAl 3438

  284 .ATGGCTCCAAGCGGA......................GTAC 266
      |||||||||||               |||:
 3438 aAlaSerProSerGlyGluAlaAspGluHisTyrGlnLeuGlnHisValA 3455

  265 AGCAA.....GGAAGTCATCAAGCGCTACACAGTAAACAGCAC...... 228
      ::||| :::|||||||||||:::|||::
 3455 snGlnLeuGluAlaSerLysThrGlyIleHisSerSerGlnArgAspGly 3471

  227 ..................ACGAATGGCTCCGATAAACATCGACGC 202
                        :::|||   :::::::::||||
 3472 AspSerAlaSerGlyProGlnValSerAsnPheThrGlnThrValAspAl 3488

  201 CATTGGCGTTGTTAGTCTTGTACAA.................ACTA 173
      |  | :::::|||| ||| |||
 3488 aProAsnSerMetGlyLeuGlyGlnLysAsnLysAlaLeuSerSerAlaValG 3505

  172 CGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGG 123
      |||||||||||||||| |||::::             |||||:::|||
 3505 lnAlaSerProThrSerProGlyGlySerProSerProSerSerSerGly 3521

  122 CAG 120
      |||
 3522 Gln 3522
```

seq_name: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR86250

CC   Analogue 4 (human FSH-beta(1-111)-linker-human-alpha(1-92)) is a
CC   specific example of a single chain gonadotropin; chimeric proteins
CC   having a chorionic gonadotropin (CG) beta-subunit at the N-terminus
CC   and a CG alpha-subunit at the C-terminus, joined by a linker of 1-16
CC   amino acids are claimed. The analogue has follicle stimulating hormone
CC   (follitropin) activity and is useful for inducing follicle development
CC   and increasing male fertility.
XX
SQ   Sequence   229 AA;

alignment_scores:
      Quality: 70.00   Length: 112
      Ratio: 1.228   Gaps: 5
Percent Similarity: 50.893   Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR86250   .

Align seg 1/1 to: AAR86250 from: 1 to: 229

```
   76 GGACACGGCTACTGTAGCTGGCACGGCATAATGGAAAACAGTCCCTGCCC 125
      |||      ||      ::||||||||     :: |||::::||:::
  118 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer........ 131

  126 TTTTGAGGGTCGTCTCCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
      :: :::                  |||||     :::::
  132 .......GlySerGlySerAlaProAspValGlnAspCysProG 146

  176 TTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCCGAGCCATTC 225
      ||||||            ::||||: |||:: |||::: |||:::|||
  146 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 162

  226 GTGTGCTGTTTACTGTGTAGCCGTTGATGACTTCCTTGCTGT........ 267
      ::: ||||||||:::||||||||||| || ||::::||:::
  163 LeuGlnCys..................MetGlyCysCysPheSerAr 172

  268 ......ACTCCGCTTGGAGCCATTGAGTGCTGTTTACCGTGCGAGC 307
      |||||||||                   ::: 
  172 gAlaTyrProThrProLeu......................Args 180

  308 ACTCGAAGACTTCCTTGCTGTAC...TCGCTGGCAGCATAGTCTGTTTGC 354
      :::||||| |||:::            :::::::::||  |||:::|||
  180 erLysLysThrMetLeuValGlnLysAsnValThrSerGluSerThrCys 196

  355 TGCCTGGTCTGGGGACACCTGTGTACTACAGTACTT 390
      |||:::|||||            |||||||||
  197 CysValAlaLysSerTyrAsnArgValThrValMet 208
```

seq_name: /SID58/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR86251

```
KW  follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS  Synthetic.
XX
FH  Key              Location/Qualifiers
FT  Peptide          1..20
FT                   /label= leader
FT  Region           21..113
FT                   /label= hCG_beta_subunit_(1-93)
FT  Misc-difference  70
FT                   /note= "Arg corresponds to CCG codon"
FT  Region           114..137
FT                   /label= hFSH_beta_subunit_(88-111)
FT  Region           138..145
FT                   /label= linker
FT  Region           146..237
FT                   /label= Gonadotropin_alpha_subunit_(1-92)
XX
PN  WO9522340-A1.
XX
PD  24-AUG-1995.
XX
PF  17-FEB-1995;   95WO-US02067.
XX
PR  18-FEB-1994;   94US-0199382.
XX
PA  (SENS-) SENSI-TEST.
XX
PI  Moyle WR;
XX
DR  WPI; 1995-302553/39.
DR  N-PSDB; AAT03227.
XX
PT  Methods for altering fertility in mammals, esp. humans - e.g.
PT  stimulating fertility by reducing the activity and/or levels of
PT  circulating glyco:protein hormones having lutropin activity
XX
PS  Example 16 and Claim 39; Fig 10; 102pp; English.
XX
CC  Analogue 5 (hCG-beta(1-93)-hFSH-beta(88-111)-linker-human alpha(1-92))
CC  is a specific example of a single chain gonadotropin having a chorionic
CC  gonadotropin (CG) beta-subunit at the N-terminus and a CG alpha-
CC  subunit at the C-terminus, joined by a linker of 1-16 amino acids.
CC  The analogue has follicle stimulating hormone (follitropin) activity
CC  but is more structurally similar to hCG then to hFSH. The analogue
CC  is useful for inducing follicle development and increasing male
CC  fertility.
XX
SQ  Sequence   237 AA;

alignment_scores:
        Quality:  70.00              Length:     112
        Ratio:     1.228             Gaps:         5
Percent Similarity: 50.893   Percent Identity:  25.000

alignment_block:
US-09-540-234-1 x AAR86251    ..

Align seq 1/1  to: AAR86251   from: 1  to: 237

  76 GGACACGGCTACTGTGTAGCTGGCACGGCATAATGGAAAAACAGTCCCTGCCC 125
     ||||||  ::||||||||||||||||||::: ||||::::||
 126 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer......... 139

 126 TTTTGGAGGGTCGTCCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
     :::  :||||||  |||||||  :::::::
 140 ......GlyGlySerGlySerAlaProAspValGlnAspCysProG 154

 176 TTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCATTC 225
     ||||||  :::||||::: |||:::  ||| ||:::::||:::
 154 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 170
```

seq_name: /SID58/gcgdata/geneseq/geneseq/AA1995.DAT:AAR86252

```
seq_documentation_block:
ID  AAR86252 standard; Protein; 237 AA.
XX
AC  AAR86252;
XX
DT  29-APR-1996  (first entry)
XX
DE  Single chain gonadotropin analogue 6.
XX
KW  Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW  alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW  inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW  follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS  Synthetic.
XX
FH  Key              Location/Qualifiers
FT  Peptide          1..20
FT                   /label= leader
FT  Region           21..120
FT                   /label= hCG_beta_subunit_(1-100)
FT  Misc-difference  70
FT                   /note= "Arg corresponds to CCG codon"
FT  Region           121..137
FT                   /label= hFSH_beta_subunit_(95-111)
FT  Region           138..145
FT                   /label= linker
FT  Region           146..237
FT                   /label= Gonadotropin_alpha_subunit_(1-92)
XX
PN  WO9522340-A1.
XX
PD  24-AUG-1995.
XX
PF  17-FEB-1995;   95WO-US02067.
XX
PR  18-FEB-1994;   94US-0199382.
XX
PA  (SENS-) SENSI-TEST.
XX
PI  Moyle WR;
XX
DR  WPI; 1995-302553/39.
DR  N-PSDB; AAT03229.
XX
PT  Methods for altering fertility in mammals, esp. humans - e.g.
PT  stimulating fertility by reducing the activity and/or levels of
PT  circulating glyco:protein hormones having lutropin activity
XX
PS  Example 17 and Claim 39; Fig 11; 102pp; English.
XX
CC  Analogue 6 (hCG-beta(1-100)-hFSH-beta(95-111)-linker-human alpha(1-92))
CC  is a specific example of a single chain gonadotropin having a chorionic
CC  gonadotropin (CG) beta-subunit at the N-terminus and a CG alpha-
CC  subunit at the C-terminus, joined by a linker of 1-16 amino acids.
```

```
 226 GTGTGCTGTTTACTGTGTAGCGCCTGATGACTTCCTTGCTGT........ 267
     ||||                         ::: ||||||
 171 LeuGlnCys.................MetGlyCysPheSerAr 180

 268 ...ACTCCCGCTTGAGCCCATTGAGTGCTGTTTACCGTCGAGC 307
     |||||||||                              :::
 180 gAlaTyrProThrProLeu.................Arg8 188

 308 ACTCGAAGACTTCCTTGCTGTAC..TCGCTGGCAGCATAGTCTGTTTGC 354
     ::::|||||| |||::  ||::::::::::::::: ||||:::||||
 188 erLysLysThrMetLeuValGlnLysAsnValThrSerGluSerThrCys 204

 355 TGCCTGGTCTGGGACACCTGTGTACTACACAGTACTT 390
     ||:::||:: :::::: ::||||||||:::
 205 CysValAlaLysSerTyrAsnArgValThrValMet 216
```

```
CC    The analogue has follicle stimulating hormone (follitropin) and
CC    luteinising hormone (lutropin) activity but is structurally more
CC    similar to hCG then to hFSH. The analogue is useful for inducing
CC    follicle development and increasing male fertility.
XX
SQ    Sequence  237 AA;

alignment_scores:
      Quality:       70.00          Length:      112
      Ratio:          1.228         Gaps:          5
Percent Similarity:  50.893         Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR86252        .

Align seg 1/1  to: AAR86252    from: 1  to: 237

 76 GGACACGGCTACTGTAGCTGGCGACGGCATAATGGAAAACAGTCCCTGCCC 125
    |||  :::||||||||||:::   ::: ||||:|||
126 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer......... 139

126 TTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
    |||    ::: ::::   |||||||    :::||||    :::::|
140 ......GlySerGlySerGlySerAlaProAspValGlnAspCysProG 154

176 TTTGTACAGAGACTAACACGCCAATGGCGTCGATGTTATCGGAGCCATTC 225
    ||||||       :::||||: ||:::|     ::: |||:|| |
154 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 170

226 GTGTGCTGTTACTGTGTAGGCGCTTGATGACTTCCTTGCTGT........ 267
    ::: |||      |||||  ||||||:
171 LeuGlnCys................MetGlyCysPheSerAr 180

268 .........ACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCGTGCAGC 307
           |||||||||
180 gAlaTyrProThrProLeu...........................ArgS 188

308 ACTCGAAGACTTCCTTGCTGTAC...TCGCTGGCAGCATAGTCTGTTTGC 354
    ::|||||| |||:::       ::::::::::::::    ||||:||||
188 erLysLysThrMetLeuValGlnLysAsnValThrSerGluSerThrCys 204

355 TGCCTGGTCTCGGGACACCTGTGTACTACAGTACTT 390
    |||:||:::   :::::  :::||||||||:::
205 CysValAlaLysSerTyrAsnArgValThrValMet 216
```

```
seq_name:   /SIDS8/gcgdata/geneseqp/geneseqp/AA2000.DAT:AAY79332

seq_documentation_block:
ID   AAY79332 standard; Protein; 467 AA.
XX
AC   AAY79332;
XX
DT   18-JUL-2000  (first entry)
XX
DE   Human EGF repeat-containing protein EGF-Hy1.
XX
KW   EGF-Hy1; epidermal growth factor; human; cancer; immunomodulator;
KW   neuroprotective; antirheumatic; antiarthritic; antiallergic;
KW   proliferative regulator; cytostatic; anticoagulant; thrombolytic;
KW   antiinflammatory; diagnosis.
XX
OS   Homo sapiens.
XX
FH   Key          Location/Qualifiers
FT   Misc-difference    38
FT                      /note= "putative initiation Met"
FT   Peptide            349..385
FT                      /note= "EGF-repeat"
FT   Peptide            387..423
FT                      /note= "EGF-repeat"
FT   Peptide            425..467
FT                      /note= "EGF-repeat"
```

```
FT   Misc-difference    226
FT                      /note= "encoded by GWA"
XX
PN   WO200017357-A1.
XX
PD   30-MAR-2000.
XX
PF   20-SEP-1999;   99WO-US21812.
XX
PR   18-SEP-1998;   98US-0157308.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Ford JE, Mulero JJ, Yeung G, Montgomery J;
XX
DR   WPI; 2000-292844/25.
DR   N-PSDB; AAZ94659.
XX
PT   Human EGF-Hy1 polypeptide and polynucleotide sequences, useful for
PT   diagnosis, prognosis and monitoring of cancer.
XX
PS   Claim 8; Page 133-135; 141pp; English.
XX
CC   The present sequence is that of human EGF-Hy1, a novel epidermal
CC   growth factor (EGF)-repeat-containing protein that bears
CC   similarity to the EGF-repeat region of Notch, a protein involved
CC   in cell fate control mechanisms regulating multicellular
CC   development. The sequence was deduced from RACE clones (see
CC   AAZ94659) isolated from a foetal brain library.  Of 18 human tissues
CC   tested, only foetal liver, foetal skin, foetal brain and adult
CC   lung expressed EGF-Hy1, suggesting expression mainly in
CC   developmental tissues.  EGF-Hy1 includes 3 EGF repeats, and
CC   polypeptides comprising these repeats are also claimed.  An
CC   additional sequence for EGF-Hy1 (see AAY79331) was also deduced
CC   from a PCR clone.  The invention also provides methods for the
CC   use of EGF-Hy1 polypeptides and polynucleotides as diagnostics,
CC   therapeutics and research reagents.  In particular, they are used
CC   for diagnosis, prognosis or monitoring of cancer.  They can also be
CC   used for stimulating and suppressing the immune system, and hence
CC   in treating immune deficiencies and disorders, including infections
CC   and autoimmune disorders such as connective tissue disease,
CC   multiple sclerosis, rheumatoid arthritis, myasthenia gravis and
CC   autoimmune inflammatory eye disease.  They can also be used to
CC   treat allergic reactions and conditions, and to regulate
CC   haematopoiesis, tissue growth, activin/inhibin,
CC   chemotaxis/chemokinesis, haemostasis, thrombolysis, receptor/ligand
CC   binding, antiinflammatory responses, leukaemia and nervous system
CC   disorders.
XX
SQ   Sequence  467 AA;

alignment_scores:
      Quality:       70.00          Length:      142
      Ratio:          1.129         Gaps:          8
Percent Similarity:  43.662         Percent Identity: 24.648

alignment_block:
US-09-540-234-1/rev x AAY79332        .

Align seg 1/1  to: AAY79332    from: 1  to: 467

402 CAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCCCCAGACCAGGCAGC 353
    |||  ||| ||| ||| |||:::||||||::::         :::
185 GlnLeuGluCysLysLeuCysProSerGlyMetTyrThrGluTyrIleHi 201

352 AAACAGACTA.........TGCTGCCAGCGAGTACAGCAAGGAAGTCTTC 312
    :::||                ||| ||| :::|||||::::::::::
201 sSerArgAsnIleSerAspCysLysLysAlaGlnCysLysGlyThr.Tyr 217

311 GAGTGCTGCGACGGTAAACAGCACTCAAATGCCTCCAAGCCGGAGTACAGCA 262
    :::|||                ||  :::|||||  |||||||||    ||
218 SerTyrSerGlyLeuGluThrCys***SerCysProLeuGlyThrTyrGl 234
```

```
                      Ratio:  1.580    Percent Identity: 57.143        Gaps:    4
Percent Similarity: 57.143

alignment_block:
US-09-540-234-1/rev x AAR13319  ..

Align seg 1/1  to: AAR13319  from: 1  to: 1023

437 AGACACGACTACATGGCTAAGAAGTAGTACCAC........TACAAAGG 397
    ::|||::     ::::::||| :||||||          |||||
812 LysHisSerSerSerGluLysThrLeuHisSerLysTyrValLysGl 828

396 GAAA...CGAAGTACTGTAGTACACAGGTGTCCCCAGACCAGGCAGCAAA 350
    |:::  |||||||::  |||::::: |||:::::::::||||::|||||
828 yArgAspArgSerCysValArgLysTyrSerGluSerArgSerSerL 845

349 CAGACATGCTGCC...AGCGAGTACAGCAAGGAAGTCTTCGAGTGCTGC 303
    |||||  |||:::   :::||| ::::
845 euAspTyrLeuSerAspSerGluGlnSerGlyGln............... 856

302 ACGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCA 253
    :::||||||  |||   :::  |||||:::
857 ........AlaThrGlnSerAlaGlnGluLysGlnGlyLysGlnGlyGlnMe 870

252 TCAAGCGCTACACAGTAAACAGCACACGAAT 222
    :::  |||::|||||||:::  |||
870 tGluArgThrHisAsnLysGlnGluLysAsn 880
```

seq_name: /SIDS8/gcgdata/geneseq/AA2000.DAT:AAY75296

```
seq_documentation_block:
ID  AAY75296 standard; Protein; 135 AA.
XX
AC  AAY75296;
XX
DT  21-MAR-2000 (first entry)
XX
DE  Neisseria meningitidis ORF 643 protein sequence SEQ ID NO:2066.
XX
KW  Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW  antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW  antibacterial; gene therapy.
XX
OS  Neisseria meningitidis.
XX
PN  WO9957280-A2.
XX
PD  11-NOV-1999.
XX
PF  30-APR-1999;   99WO-US09346.
XX
PR  01-MAY-1998;   98US-0083758.
PR  31-JUL-1998;   98US-0094869.
PR  02-SEP-1998;   98US-0098994.
PR  02-SEP-1998;   98US-0099062.
PR  09-OCT-1998;   98US-0103749.
PR  09-OCT-1998;   98US-0103794.
PR  09-OCT-1998;   98US-0103796.
PR  25-FEB-1999;   99US-0121528.
XX
PA  (CHIR ) CHIRON CORP.
PA  (GENO-) INST GENOMIC RES.
XX
PI  Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI  Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI  Tettelin H,  Venter JC;
XX
DR  WPI; 2000-062150/05.
DR  N-PSDB; AA254058.
XX
PT  Novel Neisserial polypeptides predicted to be useful antigens for
PT  vaccines and diagnostics -
XX
```

```
261 A........GGAAGTCATCAAGCGCTACACAGTAAACAGCACACGAAT. 222
    |||||||::::||||        :::|:||||:||
234 nProLysPheGlySerArgSerCysLeuSerCysProGluAsnThrSerF 251

221 .........GGCTCCGATAACATCGACGCCCATTGGCGTTGTTAGTCTT 183
    |||||    |||||::::|||
251 hrValLysArgGlyAlaValAsnIleSerAlaCysGlyVal......... 264

182 GTACAAACTACGGCTTCACCCAGCGTCTCGAGGGGCATGTTGAGACGACC 133
    |||     ::|||        |||
265 ...........ProCysProGluGlyLysPheSerArgSer.. 274

132 CTCAAAAGGCAGGGACTGTTTTCCATTATGCCGTGCCAGCTACAGTAGC 83
    :::::|||||||||::
275 ................GlyLeuMetProCysHis.........P 281

82  CGTGTCCG................GCTTGGAGGTGT 63
    |||||||
281 roCysProArgAspTyrTyrGlnProAsnAlaGlyLysAlaPheCysLeu 297

62  CTTTGTCCTTTCCTTCAGATGACC 39
    |||||||||  |||
298 AlaCysProPheTyrGlyThrThr 305
```

seq_name: /SIDS8/gcgdata/geneseq/AA1991.DAT:AAR13319

```
seq_documentation_block:
ID  AAR13319 standard; Protein; 1023 AA.
XX
AC  AAR13319;
XX
DT  22-OCT-1991 (first entry)
XX
DE  Partial Human Natural Killer receptor.
XX
KW  NK; cytotoxic drugs; tumour cell.
XX
OS  Homo sapiens.
XX
PN  US7535206-A.
XX
PD  09-JUL-1991.
XX
PF  08-JUN-1990;   90US-0143578.
XX
PR  08-JUN-1990;   90US-0535206.
XX
PA  (USSH ) NAT INST OF HEALTH.
XX
PI  Ortaldo J,  Young H,  Anderson S;
XX
DR  WPI; 1991-245694/33.
DR  N-PSDB; AAQ13114.
XX
PT  DNA encoding a natural killer cell receptor - used to develop
PT  prods. for the immuno-detection and immuno-therapy of tumours
XX
PS  Disclosure; Fig 1; 30pp; English.
XX
CC  Overlapping clones, which make up the cDNA sequence from which this
CC  sequence was deduced, were isolated from a cDNA library prepared
CC  from human NK cells purified from human peripheral blood. The
CC  purified protein can mediate the cytolytic activity of mammalian
CC  cells. It specifically distinguishes tumour cells making it a
CC  candidate for the development of products for the immunodetection
CC  and immunotherapy of tumours. See also AAQ13115.
XX
SQ  Sequence 1023 AA;

alignment_scores:
      Quality:  69.50      Length:  77
```

```
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V, Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI:  2000-062150/05.
DR   N-PSDB: AAZ54059.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 1022; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitidis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisseria bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   135 AA;

alignment_scores:
        Quality:   69.00                   Length:     122
        Ratio:      1.232                   Gaps:         8
Percent Similarity:  45.902     Percent Identity:   28.689

alignment_block:
US-09-540-234-1 x AAY75297

Align seg 1/1  to: AAY75297   from: 1  to: 135

 60 AAGACACCTCCAAGCCGGACACGG...CTACTGTAGCTGGCACGGCATAA 106
    :::::: ||||| ||||| :::|| ||| :
 25 ArgValSerProSerThrThrArgTrpMetLeu.AlaTrpSerGlyGluI 41

107 TGGAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTCAAACATGCCCCTCGA 156
    :::::  ||||||
 41 leSerAlaSerPro.............................. 45

157 GACGTCGGTGAAGCCGTAGTTGTACAGAGACTAACACGCCAATGGCCGTC 206
    ::::::::::||||||:::::::
 46 ........SerAlaAlaLeuAlaThrArgValSerLysArgThrArgAr 59
```

```
PS   Claim 2; Page 1021; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitidis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisseria bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   135 AA;

alignment_scores:
        Quality:   69.00                   Length:     122
        Ratio:      1.232                   Gaps:         8
Percent Similarity:  45.902     Percent Identity:   28.689

alignment_block:
US-09-540-234-1 x AAY75296

Align seg 1/1  to: AAY75296   from: 1  to: 135

 60 AAGACACCTCCAAGCCGGACACGG...CTACTGTAGCTGGCACGGCATAA 106
    :::::: ||||| ||||| :::|| ||| :
 25 ArgValSerProSerThrThrArgTrpMetLeu.AlaTrpSerGlyGluI 41

107 TGGAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTCAAACATGCCCCTCGA 156
    :::::  ||||||
 41 leSerAlaSerPro.............................. 45

157 GACGTCGGTGAAGCCGTAGTTGTACAGAGACTAACACGCCAATGGCCGTC 206
    :::::::::::||||||:::::::
 46 ........SerAlaAlaLeuAlaThrArgValSerLysArgThrArgAr 59

207 GATGTTATCGGAGCCATTCGTGTGCTGT..........TTACTGTGTA 244
    ::: ||| |||||||                      ::|||||||
 59 gLeuProSerAlaAlaAlaValCysCysGlyAspAlaGluIleLeuCysS 76

245 GGGCTTGATGACTTCCTTGCTACTCCGGCTTGGAGCCATTTGAGTG... 291
    |||||         ::|||:::: ||| ||
 76 erAla......ThrValSerGlyValProMetThrAlaGluMetValSer 90

292 ...........CTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    |||||     ||||||
 91 SerAlaCysArgArgArgLeuPheArgAla...........ThrSerCy 103

324 GCTGTACTCGCTGGCAGCATAGTCGTGTTTGCTGCCTGTCTCGGGACACC 373
    ::: ||| |||||     |||  ||||| :
103 sMetSerSerAlaAla........CysMetSerPheTrpGlyMetI 117

374 TGTGTACTACAGTA 387
    ::|||:::::|||
117 leCysAlaSerVal 121

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA2000.DAT:AAY75297

seq_documentation_block:
ID   AAY75297 standard; Protein; 135 AA.
XX
AC   AAY75297;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 643 protein sequence SEQ ID NO:2068.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
```

```
207 GATGTTATCGGAGCCATTCGTGTGCTGT..........TTACTGTGTA 244
    :::  |||    ||||||||             :::||||||
 59 gLeuProSerAlaAlaThrValCysGlyAspGluGluMetLeuCysS  76

245 GCGCTTGAGTGACTTCCTTCGCTGTACTCCCGTTCCGCTTGAGGCCATTGAGTG... 291
    |||||               ::::||:||:::: ||   ||
 76 erAla......ThrValSerGlyValProMetThrAlaGluMetValSer  90

292 ...........CTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
             |||||     ::: ||||||
 91 SerAlaCysArgArgArgLeuPheArgAla...........ThrSerCy 103

324 GCTGTACTCCGCTGGCAGCATAGTCTGTTTGCTGCCTGGTCTGGGGACACC 373
    :::  ||  |||||             |||  ||||||
103 sMetSerSerSerAlaAla........CysMetSerPheTrpGlyThrI 117

374 TGTGTACTACAGTA 387
    ::||||:::::|||
117 leCysAlaSerVal 121

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA1995.DAT:AAR86272

seq_documentation_block:
ID   AAR86272 standard; Protein; 229 AA.
XX
AC   AAR86272;
XX
DT   08-MAY-1996  (first entry)
XX
DE   Single chain gonadotropin analogue 4b with extra glycosylation site.
XX
KW   Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW   alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW   inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW   follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..18
FT                   /label= leader
FT   Region          19..129
FT                   /label= hFSH_beta_subunit_(1-111)
FT   Misc-difference 25
FT                   /note= "wild-type Asn at position 7 of the beta-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT   Misc-difference 42
FT                   /note= "wild-type Asn at position 24 of the beta-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT   Misc-difference 89
FT                   /note= "wild-type Asp at position 71 of the beta-
FT                   subunit is replaced by Asn to agree with
FT                   the glycosylation site motif"
FT   Misc-difference 91
FT                   /note= "wild-type Leu at position 73 of the beta-
FT                   subunit is replaced by Thr to agree with
FT                   the glycosylation site motif"
FT   Region          130..137
FT                   /label= linker
FT   Region          138..229
FT                   /label= Gonadotropin_alpha_subunit_(1-92)
FT   Misc-difference 189
FT                   /note= "wild-type Asn at position 52 of the alpha-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT   Misc-difference 215
FT                   /note= "wild-type Asn at position 78 of the alpha-
```

```
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
XX
PN   WO9522340-A1.
XX
PD   24-AUG-1995.
XX
PF   17-FEB-1995;  95WO-US02067.
XX
PR   18-FEB-1994;  94US-0199382.
XX
PA   (SENS-) SENSI-TEST.
XX
PI   Moyle WR;
XX
DR   WPI; 1995-302553/39.
XX
PT   Methods for altering fertility in mammals, esp. humans - e.g.
PT   stimulating fertility by reducing the activity and/or levels of
PT   circulating glyco:protein hormones having lutropin activity
XX
PS   Example 25; Fig 9 and Page 60; 102pp; English.
XX
CC   The single-chain gonadotropin analogue 4b (human FSH-beta(1-111)
CC   [N7X,N24X,D71N,L73T]-linker-human CG-alpha(1-92)[N52X,N78X]) is
CC   an example of a chimeric glycopeptide hormone having an extra
CC   glycosylation site. Addition of oligosaccharides has a positive
CC   effect on stability of hormones in circulation and can be used to
CC   prevent unwanted antibody or receptor interactions. The present
CC   analogue has anti-follicle stimulating hormone (follitropin)
CC   activity and is useful for treating ovarian hyperstimulation and
CC   reducing spermatogenesis.
XX
SQ   Sequence   229 AA;

alignment_scores:
         Quality:      69.00             Length:         112
         Ratio:         1.211            Gaps:             5
Percent Similarity:    50.893    Percent Identity:    25.000

alignment_block:
US-09-540-234-1 x AAR86272

Align seg 1/1  to: AAR86272     from:  1  to: 229

 76 GGACACGGCTACTGTAGCTGGCACGGCATAATGAGAAAACAGTCCCTGCCC 125
    |||    :::|||||||||||:::       :::  ||::|||
118 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer.......... 131

126 TTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
    |||          :::  :::|||||  ||||| ::::||
132 .......GlySerGlySerAlaProAspValGlnAspCysProG 146

176 TTTGTACAAGACTAACAACGGCCAATGCGTCGATGTTATCGGAGCCATTC 225
    ||||||       ||:::  ||::||   ||::::  ::::||||
146 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 162

226 GTGTGCTGTTTACTGTGTAGCGCTTGATGACTTCCTTCGCTGT........ 267
    ::: ||||||||        :::   ||||||
163 LeuGlnCys..........................MetGlyCysPheSerAr 172

268 ...........ACTCCGCTTGGAGCCATTGAGTGCTGTTACCGTGCAGC 307
             ||||||||||                 :::
172 gAlaTyrProThrProLeu.....................ArgS 180

308 ACTCGAAGACTTCCTTGCTGTAC...TCGCTGGCAGCCATAGTCTGTTTGC 354
    :::||||||         :::||||:::::||:||::|:||||
180 erLysLysThrMetLeuValGlnLys***ValThrSerGluSerThrCys 196

355 TGCCTGGTCTGGGGACACCCTGTGTACTACAGTACTT 390
    ||::::::    :::::::    ::||||||||:::
```

197 CysValAlaLysSerTyrAsnArgValThrValMet 208
seq_name: /SIDS8/gcgdata/geneseq/AA1995.DAT:AAR86262

seq_documentation_block:
ID   AAR86262 standard; Protein; 229 AA.
XX
AC   AAR86262;
XX
DT   08-MAY-1996 (first entry)
XX
DE   Partially deglycosylated single chain gonadotropin analogue 4a.
XX
KW   Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW   alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW   inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW   follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..18
FT   /label= leader
FT   Region   19..129
FT   /label= hFSH_beta_subunit_(1-111)
FT   Misc-difference 25
FT   /note= "wild-type Asn at position 7 of the beta-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
FT   Misc-difference 42
FT   /note= "wild-type Asn at position 24 of the beta-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
FT   Region   130..137
FT   /label= linker
FT   Region   138..229
FT   /label= Gonadotropin_alpha_subunit_(1-92)
FT   Misc-difference 189
FT   /note= "wild-type Asn at position 52 of the alpha-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
FT   Misc-difference 215
FT   /note= "wild-type Asn at position 78 of the alpha-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
XX
PN   WO9522340-A1.
XX
PD   24-AUG-1995.
XX
PF   17-FEB-1995; 95WO-US02067.
XX
PR   18-FEB-1994; 94US-0199382.
XX
PA   (SENS-) SENSI-TEST.
XX
PI   Moyle WR;
XX
DR   WPI; 1995-302553/39.
XX
PT   Methods for altering fertility in mammals, esp. humans - e.g.
PT   stimulating fertility by reducing the activity and/or levels of
PT   circulating glyco:protein hormones having lutropin activity
XX
PS   Example 24; Fig 9 and Page 60; Page 102pp; English.
XX
CC   The single-chain gonadotropin analogue 4a (human FSH-beta(1-111)
CC   [N7X,N24X]-linker-human CG-alpha(1-92)[N52X,N78X]) can be derived
CC   from analogue 4 by removing at least one of the four glycosylation

CC   sites. The partially deglycosylated analogue has anti-follicle
CC   stimulating hormone (follitropin) activity and is useful for
CC   treating ovarian hyperstimulation and reducing spermatogenesis.
XX
SQ   Sequence 229 AA;

alignment_scores:
        Quality: 69.00   Length: 112
        Ratio: 1.211   Gaps: 5
Percent Similarity: 50.893   Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR86262

Align seg 1/1 to: AAR86262   from: 1   to: 229

 76 GGACACGCTACTGTAGCTGGCACGGCATAATGAAAAACAGTCCCTGCCC 125
    ||| :::|||||||||::   ::: ||||:::|||
118 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer......... 131

126 TTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
     ||| ::: :: |||||| :::::
132 ....GlySerGlySerGlySerAlaProAspValGlnAspCysProG 146

176 TTTGTACAGACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCATTC 225
     |||||| :::||::: ||||:: ::: |||::: ||:::
146 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 162

226 GTGTGCGTGTTTACTGTGTAGCGCTTGATGACTTCCTTGCTGT......... 267
     ::: ||| ||||||
163 LeuGlnCys..................MetGlyCysCysPheSerAr 172

268 ........ACTCCGCTTGGAGCCATTTGAGTGCGTGTTTACCGTGCAGC 307
          |||||||| :::
172 gAlaTyrProThrProLeu.........................ArgS 180

308 ACTCGAAGACTTCCTTGCTGTAC...TCGCTGGCAGCATAGTCGTTTGC 354
     ::::|||||| |||::: ||:::::::::||| |||::|||
180 erLysLysThrMetLeuValGlnLys***ValThrSerGluSerThrCys 196

355 TGCCTGGTCTGGGGACACCTGTGTACTACAGTACTT 390
     ||::::||: ::: :::||||||::||||||
197 CysValAlaLysSerTyrAsnArgValThrValMet 208

seq_name: /SIDS8/gcgdata/geneseq/AA1995.DAT:AAR86273

seq_documentation_block:
ID   AAR86273 standard; Protein; 237 AA.
XX
AC   AAR86273;
XX
DT   13-MAY-1996 (first entry)
XX
DE   Single chain gonadotropin analogue 5b with extra glycosylation site.
XX
KW   Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW   alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW   inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW   follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..20
FT   /label= leader
FT   Region   21..113
FT   /label= hCG_beta_subunit_(1-93)
FT   Misc-difference 33
FT   /note= "wild-type Asn at position 13 of the beta-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"

Align seg 1/1 to: AAR86273   from: 1 to: 237

```
          76  GGACACGGCTACTGTAGCTGGCACGGCATAATGAGAAAACAGTCCCTGCCC  125
              |||   ::::||||||||||||||||||         ::  |||::|||||
         126  GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer........  139

         126  TTTTGAGGGTCGTCTCAAACATGCCCTCGAGACGCGTCGGTGAAGCCCTAG  175
              ||||       :: :: |||||        |||||  :::::||
         140  ......GlySerGlySerAlaProAspValGlnAspCysProGly  154

         176  TTTGTACAAGACTAACACGCCAATGGCGTCGCATGTTATCGGAGCCATTC  225
              ||||||              ::|||::  |||::: :|||:::  |||::
         154  luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle  170

         226  GTGTGCTGTTTACTGTGTAGCGCTTGATGACTTCCTGCTGT.........  267
              ::: |||   |||                  ::: |||||
         171  LeuGlnCys..............MetGlyCysPheSerAr  180

         268  ..........ACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCGTGCAGC  307
              |||||||||                              ::
         180  gAlaTyrProThrProLeu.........................ArgS  188

         308  ACTCGAAGACTTCCTTGCTGTAC...TCGCTGCAGCATAGTCTGTTTGC  354
              ::|||||| ||:::         :::::::::::|||  |||:::|||
         188  erLysLysThrMetLeuValGlnLys***ValThrSerGluSerThrCys  204

         355  TGCCTGGTCTGGGGACACCTGTGTACTACAGTACTT  390
              |||:|||||     ::::::   ::|||||||::
         205  CysValAlaLysSerTyrAsnArgValThrValMet  216
```

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA1995.DAT:AAR86274

```
seq_documentation_block:
ID  AAR86274 standard; Protein; 237 AA.
XX
AC  AAR86274;
XX
DT  13-MAY-1996 (first entry)
XX
DE  Single chain gonadotropin analogue 6b with extra glycosylation site.
XX
KW  Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW  alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW  inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW  follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS  Synthetic.
XX
FH  Key              Location/Qualifiers
FT  Peptide          1..20
FT                   /label= leader
FT  Region           21..120
FT                   /label= hCG_beta_subunit_(1-100)
FT  Misc-difference  33
FT                   /note= "wild-type Asn at position 13 of the beta-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT  Misc-difference  50
FT                   /note= "wild-type Asn at position 30 of the beta-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT  Misc-difference  70
FT                   /note= "Arg corresponds to CCG codon"
FT  Misc-difference  98
FT                   /note= "wild-type Pro at position 78 of the beta-
FT                   subunit is replaced by another amino acid
FT                   to agree with the glycosylation site motif"
FT  Misc-difference  99
FT                   /note= "wild-type Val at position 79 of the beta-
FT                   subunit is replaced by Thr to agree with the
FT                   glycosylation site motif"
```

Right/second column:

```
FT  Misc-difference  50
FT                   /note= "wild-type Asn at position 30 of the beta-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT  Misc-difference  70
FT                   /note= "Arg corresponds to CCG codon"
FT  Misc-difference  98
FT                   /note= "wild-type Pro at position 78 of the beta-
FT                   subunit is replaced by another amino acid
FT                   to agree with the glycosylation site motif"
FT  Misc-difference  99
FT                   /note= "wild-type Val at position 79 of the beta-
FT                   subunit is replaced by Thr to agree with the
FT                   glycosylation site motif"
FT  Region           114..137
FT                   /label= hFSH_beta_subunit_(88-111)
FT  Region           138..145
FT                   /label= linker
FT  Region           146..237
FT                   /label= Gonadotropin_alpha_subunit_(1-92)
FT  Misc-difference  197
FT                   /note= "wild-type Asn at position 52 of the alpha-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
FT  Misc-difference  223
FT                   /note= "wild-type Asn at position 78 of the alpha-
FT                   subunit is pref. replaced by another amino
FT                   acid (esp. Gln) to remove a glycosylation
FT                   site"
XX
PN  WO9522340-A1.
XX
PD  24-AUG-1995.
XX
PF  17-FEB-1995;  95WO-US02067.
XX
PR  18-FEB-1994;  94US-0199382.
XX
PA  (SENS-) SENSI-TEST.
XX
PI  Moyle WR;
XX
DR  WPI; 1995-302553/39.
XX
PT  Methods for altering fertility in mammals, esp. humans - e.g.
PT  stimulating fertility by reducing the activity and/or levels of
PT  circulating glyco;protein hormones having lutropin activity
XX
PS  Example 25; Fig 10 and Page 60; 102pp; English.
XX
CC  The single-chain gonadotropin analogue 5b (human CG-beta(1-93)
CC  [N13X,N30X,P78X,V79T]-hFSH-beta(88-111)-linker-human CG-alpha(1-92)
CC  [N52X,N78X]) is an example of a chimeric glycopeptide hormone having
CC  an extra glycosylation site. Addition of oligosaccharides has a
CC  positive effect on stability of hormones in circulation and can be
CC  used to prevent unwanted antibody or receptor interactions. The
CC  present analogue has anti-follicle stimulating hormone (follitropin)
CC  activity and is useful for treating ovarian hyperstimulation and
CC  reducing spermatogenesis.
XX
SQ  Sequence  237 AA;
```

alignment_scores:
```
  Quality:             69.00        Length:      112
    Ratio:             1.211        Gaps:          5
Percent Similarity:   50.893   Percent Identity:  25.000
```

alignment_block:
US-09-540-234-1 x AAR86273

```
FT   Region              121..137
FT                       /label= hFSH_beta_subunit_(95-111)
FT   Region              138..145
FT                       /label= linker
FT   Region              146..237
FT                       /label= Gonadotropin_alpha_subunit_(1-92)
FT   Misc-difference     197
FT                       /note= "wild-type Asn at position 52 of the alpha-
FT                       subunit is pref. replaced by another amino
FT                       acid (esp. Gln) to remove a glycosylation
FT                       site"
FT   Misc-difference     223
FT                       /note= "wild-type Asn at position 78 of the alpha-
FT                       subunit is pref. replaced by another amino
FT                       acid (esp. Gln) to remove a glycosylation
FT                       site"
XX
PN   WO9522340-A1.
XX
PD   24-AUG-1995.
XX
PF   17-FEB-1995;  95WO-US02067.
XX
PR   18-FEB-1994;  94US-0199382.
XX
PA   (SENS-) SENSI-TEST.
XX
PI   Moyle WR;
XX
DR   WPI; 1995-302553/39.
XX
PT   Methods for altering fertility in mammals, esp. humans - e.g.
PT   stimulating fertility by reducing the activity and/or levels of
PT   circulating glyco:protein hormones having lutropin activity
XX
PS   Example 25; Fig 11 and Page 60; 102pp; English.
XX
CC   The single-chain gonadotropin analogue 6b (human CG-beta(1-100)
CC   [N13X,N30X,P78X,V79T]-hFSH-beta(95-111)-linker-human CG-alpha(1-92)
CC   [N52X,N78X]) is an example of a chimeric glycopeptide hormone having
CC   an extra glycosylation site. Addition of oligosaccharides has a
CC   positive effect on stability of hormones in circulation and can be
CC   used to prevent unwanted antibody or receptor interactions. The
CC   present analogue has anti-follicle stimulating hormone (follitropin)
CC   and anti-luteinising hormone (lutropin) activity and is useful for
CC   treating ovarian hyperstimulation and reducing spermatogenesis.
XX
SQ   Sequence    237 AA;


alignment_scores:
     Quality:     69.00              Length:       112
     Ratio:       1.211              Gaps:           5
     Percent Similarity:  50.893     Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR86274

Align seg 1/1  to: AAR86274  from: 1  to: 237

  76 GGACACGGCTACTGTAGCTGGCACGGCATAATGGAAAACAGTCCCTGCCC 125
     ||| :::||||||||||::         ::  |||:::|||
 126 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer......... 139

 126 TTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
     |||                  |||||:|| ||||||
 140 ......GlySerGlySerGlySerAlaProAspValGlnAspCysProG 154

 176 TTTGTACAAGACTAACACGCCAATGGCGTCGCGTCGCATGTTATCGGAGCCATTC 225
     ||||||      :::||||::::  :::|||::::  ::: :::|||:::
 154 luCysThrLeuGlnGluAsnProPheSerGlnProGlyAlaProIle 170

 226 GTGTGCTGCTGTTTACTGTGTAGCCGTTGAGCCTTGATGACTTCCTTGCTGT...... 267
```

```
171 LeuGlnCys..............................MetGlyCysPheSerAr 180
    :::    |||                                  ||||||
268 .........ACTCCGCTTGGAGCCATTTGAGTGCTGTGTTACCGTGCAGC 307
    ||||||||||                                   :::
180 gAlaTyrProThrProLeu........................ArgS 188
    ||||||||||
308 ACTCGAAGACTTCCTTGCTGTAC...TCGCTGGCAGCATAGTCTGTTTGC 354
    :::||||| |||::                 :|| :::::::::  |||:||||
188 erLysLysThrMetLeuValGlnLys***ValThrSerGluSerThrCys 204
    |||::: |||:::                      ::||||||
355 TGCCTGGCTGTGGGGACACCCTGTGTACTACAGTACTT 390
    ||| ::::::  ::::    :::::  :::||||||
205 CysValAlaLysSerTyrAsnArgValThrValMet 216

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA1995.DAT:AAR86263

seq_documentation_block:
ID   AAR86263 standard; Protein; 237 AA.
XX
AC   AAR86263;
XX
DT   08-MAY-1996  (first entry)
XX
DE   Partially deglycosylated single chain gonadotropin analogue 5a.
XX
KW   Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW   alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW   inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW   follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS   Synthetic.
XX
FH   Key                 Location/Qualifiers
FT   Peptide             1..20
FT                       /label= leader
FT   Region              21..113
FT                       /label= hCG_beta_subunit_(1-93)
FT   Misc-difference     33
FT                       /note= "wild-type Asn at position 13 of the beta-
FT                       subunit is pref. replaced by another amino
FT                       acid (esp. Gln) to remove a glycosylation
FT                       site"
FT   Misc-difference     50
FT                       /note= "wild-type Asn at position 30 of the beta-
FT                       subunit is pref. replaced by another amino
FT                       acid (esp. Gln) to remove a glycosylation
FT                       site"
FT   Misc-difference     70
FT                       /note= "Arg corresponds to CCG codon"
FT   Region              114..137
FT                       /label= hFSH_beta_subunit_(88-111)
FT   Region              138..145
FT                       /label= linker
FT   Region              146..237
FT                       /label= Gonadotropin_alpha_subunit_(1-92)
FT   Misc-difference     197
FT                       /note= "wild-type Asn at position 52 of the alpha-
FT                       subunit is pref. replaced by another amino
FT                       acid (esp. Gln) to remove a glycosylation
FT                       site"
FT   Misc-difference     223
FT                       /note= "wild-type Asn at position 78 of the alpha-
FT                       subunit is pref. replaced by another amino
FT                       acid (esp. Gln) to remove a glycosylation
FT                       site"
XX
```

PR   18-FEB-1994;   94US-0199382.
XX
PA   (SENS-) SENSI-TEST.
XX
PI   Moyle WR;
XX
DR   WPI; 1995-302553/39.
XX
PT   Methods for altering fertility in mammals, esp. humans - e.g.
PT   stimulating fertility by reducing the activity and/or levels of
PT   circulating glyco:protein hormones having lutropin activity
XX
PS   Example 24; Fig 10 and Page 60; 102pp; English.
XX
CC   The single-chain gonadotropin analogue 5a (human CG-beta(1-93)
CC   [N13X,N30X]-hFSH-beta(88-111)-linker-human CG-alpha(1-92)[N52X,N78X])
CC   can be derived from analogue 5 by removing at least one of the four
CC   glycosylation sites. The partially deglycosylated analogue has anti-
CC   follicle stimulating hormone (follitropin) activity and is useful for
CC   treating ovarian hyperstimulation and reducing spermatogenesis.
XX
SQ   Sequence   237 AA;

alignment_scores:
     Quality: 69.00   Length: 112
     Ratio: 1.211   Gaps: 5
Percent Similarity: 50.893   Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR86263

Align seg 1/1 to: AAR86263   from: 1   to: 237   .

```
  76 GGACACGGCTACTGTAGCTGGCACGGCCATAATGGAAAAACAGTCCCTGCCC 125
     ||| :::|||||||||::::       |||  ||  :::||
 126 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer........ 139

 126 TTTTGAGGGTGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
     |||     ::: ||| |||||||     ||||||
 140 .......GlySerGlySerGlySerAlaProAspValGlnAspCysProG 154

 176 TTTGTACAGAGACTAACAACGCACCAATGGCGTCGATGTTATCGGAGCCATTC 225
     ||||||        :::|||::: |||::  ::: |||::|
 154 luCysThrLeuGlnGluAsnProPhePheSerGlnProGlyAlaProIle 170

 226 GTGTGCGTGTTTACTGTGTAGCGCGTTGATGACTTCCTTGCTGT........ 267
     ::: |||            ::: ||||||||||||
 171 LeuGlnCys..............MetGlyCysCysPheSerAr 180

 268 .........ACTCCGCTTGGAGCCATTGAGTGCTGTTTACCGTGCGAGC 307
            ||||||||
 180 gAlaTyrProThrProLeu...................ArgS 188

 308 ACTCGAAGACTTCCTTGCTGTAC...TCGCTGGCAGCCATAGTCTGTTTGC 354
     ::|||||| |||::::       ::::::::::: ||::::|||
 188 erLysLysThrMetLeuValGlnLys***ValThrSerGluSerThrCys 204

 355 TGCCTGGTCTGGGGACACCTGTGTTACTACAGTACTT 390
     :::::::::   ::::::::  ||||||||||||
 205 CysValAlaLysSerTyrAsnArgValThrValMet 216
```

seq_name: /SIDS8/gcgdata/geneseq/AA1995.DAT:AAR86264

seq_documentation_block:
ID   AAR86264 standard; Protein; 237 AA.
XX
AC   AAR86264;
XX
DT   08-MAY-1996   (first entry)
XX
DE   Partially deglycosylated single chain gonadotropin analogue 6a.
XX

KW   Single chain gonadotropin; human chorionic gonadotropin; hCG;
KW   alpha; beta; subunit; analogue; glycoprotein hormone; fertility;
KW   inhibit; stimulate; increase; lutropin; luteinising hormone; LH;
KW   follicle stimulating hormone; FSH; vaccine; contraceptive.
XX
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..20
FT   /label= leader
FT   Region   21..120
FT   /label= hCG_beta_subunit_(1-100)
FT   Misc-difference 33
FT   /note= "wild-type Asn at position 13 of the beta-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
FT   Misc-difference 50
FT   /note= "wild-type Asn at position 30 of the beta-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
FT   Misc-difference 70
FT   /note= "Arg corresponds to CCG codon"
FT   Region   121..137
FT   /label= hFSH_beta_subunit_(95-111)
FT   Region   138..145
FT   /label= linker
FT   Region   146..237
FT   /label= Gonadotropin_alpha_subunit_(1-92)
FT   Misc-difference 197
FT   /note= "wild-type Asn at position 52 of the alpha-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
FT   Misc-difference 223
FT   /note= "wild-type Asn at position 78 of the alpha-
FT   subunit is pref. replaced by another amino
FT   acid (esp. Gln) to remove a glycosylation
FT   site"
XX
PN   WO9522340-A1.
XX
PD   24-AUG-1995.
XX
PF   17-FEB-1995;   95WO-US02067.
XX
PR   18-FEB-1994;   94US-0199382.
XX
PA   (SENS-) SENSI-TEST.
XX
PI   Moyle WR;
XX
DR   WPI; 1995-302553/39.
XX
PT   Methods for altering fertility in mammals, esp. humans - e.g.
PT   stimulating fertility by reducing the activity and/or levels of
PT   circulating glyco:protein hormones having lutropin activity
XX
PS   Example 24; Fig 11 and Page 60; 102pp; English.
XX
CC   The single-chain gonadotropin analogue 6a (human CG-beta(1-100)
CC   [N13X,N30X]-hFSH-beta(95-111)-linker-human CG-alpha(1-92)[N52X,N78X])
CC   can be derived from analogue 6 by removing at least one of the four
CC   glycosylation sites. The partially deglycosylated analogue has anti-
CC   follicle stimulating hormone (follitropin) and anti-luteinising
CC   hormone (lutropin) activity and is useful for treating ovarian
CC   hyperstimulation and reducing spermatogenesis.
XX
SQ   Sequence   237 AA;

alignment_scores:

CC   The invention provides novel polypeptides derived from the C-terminal
CC   region of human and mouse agouti-related transcript (ART) protein. These
CC   peptides can be used to inhibit binding of melanocyte stimulating
CC   hormones (MSHs) to melanocortin receptors. The ART proteins form a part
CC   of fusion proteins which have an amino acid sequence from the ART
CC   protein fused at its carboxy terminus to one or more amino acid sequences
CC   not derived from the ART protein. The ART polypeptides can be used to
CC   identify inhibitors and allosteric enhancers of the binding of the ART
CC   polypeptide to melanocortin receptors. The ART protein is also a
CC   regulator of human obesity, and substances that potentiate the effect of
CC   the ART protein on melanocortin receptors are likely to be of value in
CC   the control of body weight. Sequences AAY21861-863 represent specific
CC   examples of ART polypeptides that can be used in the invention. The
CC   present sequence represents a fusion protein containing the ART protein.
XX
SQ   Sequence   684 AA;

alignment_scores:
        Quality: 69.00   Length: 87
        Ratio: 2.029   Gaps: 4
Percent Similarity: 39.080   Percent Identity: 24.138

alignment_block:
US-09-540-234-1 x AAY21873   .

Align seg 1/1   to: AAY21873   from: 1   to: 684

194 CGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTACTGTGT 243
    |||::: ||| ||||||||| ||| |||||||||::|| |||:::
 82 ArgSerArgArgCysValArgLeuHisGluSer........CysLe 95

244 AGCGCTTGATGACTTCCTTGCTGTACTCCGCTTGGAGCCATTGAGTGCT 293
    :::   ::: |||
 95 uGly................................GlnGlnValProCysC 102

294 GTTTACCGTGCAGCACTCGAAGACTTCCTTGCTGTACTCCCTGGCAGCAT 343
    || ||||||:::
102 ysAspProCysAla.................................. 106

344 AGTCTGTTTGCTGCCTGGTCTCGGGACACCTGTACTACAGTACTTCGT 393
                           |||||| :::||:||
107 ....ThrCysTyrCysArgPhePheAs 114

394 TTCCCTTGT.............AGTGGTACTACTTCTTAGCCATGTACTC 431
    ::||:|||| |||||||||:::|
114 nAlaPheCysTyrCysArgLysLeuGlyThrAlaMetAsnProCysSerA 131

432 GTGTCGTTCG 442
    ||| |||
131 rgThrMetSer 134

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA2000.DAT:AAG03788

seq_documentation_block:
ID   AAG03788 standard; Protein; 131 AA.
XX
AC   AAG03788;
XX
DT   06-OCT-2000   (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 7869.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.

        Quality: 69.00   Length: 112
        Ratio: 1.211   Gaps: 5
Percent Similarity: 50.893   Percent Identity: 25.000

alignment_block:
US-09-540-234-1 x AAR86264   .

Align seg 1/1   to: AAR86264   from: 1   to: 237

 76 GGACACGGCTACTGTAGCTGGCACGGCATAATGGAAAACAGTCCCTGCCC 125
    ||| ::||||||||||||:: :|| |||:::|||
 76 GlyProSerTyrCysSerPheGlyGluMetLysGluGlySer........ 139

126 TTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAG 175
    || ||| ::| |||||| :::::::
126 ....GlySerGlySerGlySerAlaProAspValGlnAspCysProG 154

176 TTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCATTC 225
    |||||| :::|||::|| :: ||||||
154 luCysThrLeuGlnGluAsnProPhePheSerGlnProGlyAlaProIle 170

226 GTGTGCTGTTACTGTGTAGCGCTTGATGACTTCCTTGCTGT........ 267
    |:: ||| ::| |||||||
171 LeuGlnCys................................MetGlyCysSerPheSerAr 180

268 .........ACTCCGTTGGAGCCATTGAGTGCTGTTTACCGTGCAGC 307
          ||||||| :::
180 gAlaTyrProThrProLeu......................ArgS 188

308 ACTCGAAGACTTCCTTGCTGTAC..TCGCTGGCAGCATAGTCTGTTTGC 354
    ::|||||| ||::: ||:::::::::|||| |||:||||
188 erLysThrMetLeuValGlnLys***ValThrSerGluSerThrCys 204

355 TGCCTGGTCTCGGGACACCTGTGTACTACAGTACTT 390
    |||:::   :::::||:||||:::
205 CysValAlaLysSerTyrAsnArgValThrValMet 216

seq_name: /SIDS8/gcgdata/geneseqp/geneseqp/AA1999.DAT:AAY21873

seq_documentation_block:
ID   AAY21873 standard; protein; 684 AA.
XX
AC   AAY21873;
XX
DT   20-SEP-1999   (first entry)
XX
DE   Amino acid sequence of fusion protein containing ART protein.
XX
KW   ART; agouti-related transcript; melanocyte stimulating hormone; MSH;
KW   melanocortin receptor; allosteric enhancer; obesity regulator.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9931508-A1.
XX
PD   24-JUN-1999.
XX
PF   11-DEC-1998;   98WO-US26457.
PR   16-DEC-1997;   97US-0069747.
XX
PA   (MERI) MERCK & CO INC.
XX
PI   Fong TM, Tota MR, Van Der Ploeg LHT;
XX
DR   WPI; 1999-430057/36.
XX
PT   Inhibiting binding of melanocyte stimulating hormones to
PT   melanocortin receptors
XX
PS   Claim 2; Page 15; 53pp; English.
XX

```
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC03794.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 13; SEQ ID 7869; 71pp + CD-ROM; English.
XX
CC   The present sequence is a polypeptide encoded by one of a large number
CC   of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC   were prepared from total human RNAs or polyA+ RNAs derived from 30
CC   different tissues. EST sequences usually correspond mainly to the 3'
CC   untranslated region (UTR) of the mRNA because they are often obtained
CC   from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC   isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC   those cases where longer cDNA sequences have been obtained, the full 5'
CC   UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC   ends and can therefore be used to obtain full length cDNAs and genomic
CC   DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC   chromosome mapping procedures. They are used to obtain upstream
CC   regulatory sequences and to design expression and secretion vectors.
XX
SQ   Sequence   131 AA;

alignment_scores:
     Quality:         68.50          Length:          109
     Ratio:           1.522          Gaps:              6
Percent Similarity:   41.284    Percent Identity:   23.853

alignment_block:
US-09-540-234-1 x AAG03788       ..

Align seg 1/1  to: AAG03788   from: 1  to: 131

  94 TGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTCAA 143
                   ::::::::::: :::::::::::|||||||||
  53 TrpHisGlyLeuThrArgArgSerPro.................... 61

 144 ACATGCCCCCCGGAGAGCGTCGGTGAGCGCGTAGTTTGTACAAGACTAACAA 193
                   |||:: :::|||  ||| :::|||
  62 .........AspLeuTrpValAlaLysIleCysGlyArgSerValV 74

 194 CGGCAATGGCGCTGCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGTGT 243
     :: ::||||  :::  ::||||||| :::
  74 al.SerTrpMetGlySerHis...AsnHisSerLeuProLeuAlaGl 89

 244 AGCGGCTTGATGACTTCCTTGCTGACTCCCGCTTGGAGCCATTTGAGTGCT 293
     :::||||||||  |||  ::|||
  89 yGlyGlyAlaPhePheLeuAla.................... 96

 294 GTTTACCGTGCCAGCAGCTGCGAAGACTTCCTTGCTGTACTCCCTGCCAGCAT 343
                   |||||||||
  97 .............. ProCysCys.................... 99

 344 AGTCTGTTTGCTGCCTCTGGGGACACCTGTACTACAGATACTTCGT 393
       ::||||||||||  |||  ::||||
 100 .........SerTrpAlaGlyAsp...CysProThrProPhePh 111

 394 TTCCCTTTGTAGTGGTACTACTTCT 418
     ||||||||||||
 111 eIleLeuCysGlySer******Ser 119

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:AAB56069
```

```
seq_documentation_block:
ID   AAB56069 standard; Protein; 296 AA.
XX
AC   AAB56069;
XX
DT   08-MAR-2001 (first entry)
XX
DE   Skin cell protein, SEQ ID NO: 458.
XX
KW   Mouse; skin cell; cytostatic; antiinflammatory; anti-HIV;
KW   nootropic; neuroprotective; vulnery; immunomodulatory; vaccine;
KW   keratinocyte growth stimulation; cancer; angiogenesis inhibition;
KW   inflammation; neurological disease.
XX
OS   Mus sp.
XX
XX
PN   WO200069884-A2.
XX
PD   23-NOV-2000.
XX
PF   15-MAY-2000; 2000WO-NZ00075.
XX
PR   14-MAY-1999;   99US-0312283.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Watson JD, Strachan L, Onrust R,   Sleeman M,  Kumble KD,  Murison JG;
XX
DR   WPI; 2001-007495/01.
DR   N-PSDB; AAC99790.
XX
PT   New isolated polynucleotide used in the identification of genetic
PT   disorders and encoding polypeptides used for treating inflammatory
PT   disease, cancer and neurological diseases -
XX
PS   Claim 4; Page 347; 352pp; English.
XX
CC   The present sequence is a polypeptide which is expressed in
CC   mammalian skin cells. The polypeptide is useful for stimulating
CC   keratinocyte growth and motility, inhibiting the growth of cancer cells,
CC   modulating angiogenesis, inhibiting angiogenesis and vascularisation of
CC   tumours, modulating skin inflammation, stimulating the growth of
CC   epithelial cells, inhibiting the binding of human immunodeficiency virus
CC   (HIV)-1 to leukocytes, and treating inflammatory disease, cancer and
CC   neurological diseases. The polynucleotide can be used as a marker, in
CC   the identification of genetic disorders, and for the design of
CC   oligonucleotides for examining expression patterns.
XX
SQ   Sequence   296 AA;

alignment_scores:
     Quality:        68.50           Length:          162
     Ratio:          1.123           Gaps:              8
Percent Similarity:  37.654    Percent Identity:   24.074

alignment_block:
US-09-540-234-1/rev x AAB56069       ..

Align seg 1/1  to: AAB56069   from: 1  to: 296

 371 TGTCCCCAGACCAGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA 322
     ||||||||| |||            :: |||
  18 CysProGlnGlyArgPheGlyProSerCysAlaHisValCysThrCysGl 34

 321 GGAAGTCTTCGAGTGCTGCACGTAAACACTCAAATGGTCTCCAAGCG 272
     :::                  |||:::|||   :::::|||
  34 yGlnGlyAlaAlaCysAspCysProValSerGlyThrCysIleCysProProG 51

 271 GA.............GTACAGCAAGGAAGTCATCAAGCGCTACACAGT 237
     ||                        :::
  51 lyLysThrGlyHisCysGluArgGlyCysProGlnAspArgPheGly 67
```

```
236 AAA......CAGCAC.................................... 228
        |||  ::|||
 68 LysGlyCysGluHisLysCysAlaCysArgAsnGlyGlyLeuCysHisAl 84

227 .ACGAATGGCTCCGATAAC................................ 210
     |||||||||||||  :::
 84 aThrAsnGlySerCysProLeuGlyTrpMetGlyProHisCysG 101

209 ............ATCGACGCCATTGGCGTGTTAGTCTTGTACAAACT 174
                ||:::::   |||::::::
101 luHisAlaCysProAlaGlyTyrGlyAlaCysLeuLeuGluCys 117

173 ................ACGGCTTCACCGACGCTCCGAGGGGCATGTTT 142
                    :::::::||||||||  ||||||||||
118 SerCysGlnAsnAsnGlySerCysGluProThrSer...GlyAlaCysLe 133

141 GAGAGACCCTCA...................................... 129
        | |||::::||

133 uCysGlyProGlyPheTyrGlyGlnAlaCysGluAspThrCysProAlaG 150

128 ....AAAGGGCAGGGACTGTTTCCATTATGCCGTGCCAGCTACAGTAG 84
        ||:::||||             ::  ||||||
150 lyPheHisGlySerGlyValCysGlnArgValCysGlnGlnGlyAla 166

 83 CCGTGTCCG......GCTTGGAGGTGTCTTTGTCCT 54
        |||||  :::  ||||||||||||
167 ProCysAspProValSerGlyArgCysLeuCysPro 178
```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY76014

seq_documentation_block:
ID   AAY76014 standard; Protein; 299 AA.
XX
AC   AAY76014;
XX
DT   27-MAR-2000 (first entry)
XX
DE   Murine EGF family protein, SEQ ID NO:192.
XX
KW   Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;
KW   embryonic skin cell; keratinocyte stem cell; transit amplifying cell;
KW   secreted; transmembrane; inflammation; cancer; neurological disease;
KW   angiogenesis; tumour vascularisation; growth disorder;
KW   developmental disorder; skin wound; hair follicle disorder;
KW   anti-inflammatory; cytostatic; neuroprotective; vulnery.
XX
OS   Mus sp.
XX
PN   WO9955865-A1.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999; 99WO-NZ000051.
XX
PR   29-APR-1998; 98US-0069726.
PR   09-NOV-1998; 98US-0188930.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murison JG;
XX
DR   WPI; 2000-072177/06.
DR   N-PSDB; AA261678.
XX
PT   Novel polynucleotides useful for the treatment of various conditions
PT   including wounds and cancer -
XX
PS   Claim 4; Page 132; 235pp; English.
XX
CC   The invention relates to novel nucleic acid sequences derived from rat
CC   dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,
CC   and mouse embryonic skin, keratinocyte stem cells and transit amplifying
CC   cells. Polypeptides of the invention may be used to treat inflammation,
CC   cancer and neurological diseases. The proteins may be used to stimulate
CC   the growth and motility of keratinocytes, to inhibit the growth of
CC   cancer cells, to modulate angiogenesis and tumour vascularisation, to
CC   modulate skin inflammation. The invention may also be used
CC   inhibit binding of HIV-1 to leukocytes. The invention may also be used
CC   to treat growth and developmental defects, skin wounds and hair follicle
CC   disorders. Sequences AAY75942-Y76123 represent polypeptides encoded
CC   by cDNA sequences derived from several mouse, rat or human skin cell
CC   types. Sequences AAY75942-Y75947, AAY76020-Y76021, AAY76094-Y76104 and
CC   AAY76119 are proteins with an N-terminal signal sequence, indicating
CC   that they are secreted. Sequences AAY75986-Y75989, AAY76061-Y76071,
CC   AAY76106-Y76109 and AAY76121-Y76122 are proteins with one or more
CC   putative transmembrane domains.
XX
SQ   Sequence   299 AA;

alignment_scores:
        Quality: 68.50   Length: 162
        Ratio: 1.123   Gaps: 8
Percent Similarity: 37.654   Percent Identity: 24.074

alignment_block:
US-09-540-234-1/rev x AAY76014
Align seg 1/1 to: AAY76014 from: 1 to: 299

```
371 TGTCCCCAGACCAGGCAGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA 322
        |||||||| |||            ||      :::     ||
 21 CysProGlnHisGlyArgPheGlyProSerCysAlaHisValCysThrCysGl 37

321 GGAAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG 272
        :::          ||          :::::::|||   :::||| |
 37 yGlnGlyAlaAlaCysAspProValSerGlyThrCysIleCysProProG 54

271 GA..........GTACAGCAAGGAAGTCATCAAGCGCTACACACAGT 237
        ||                      ||::::||||   |||
 54 lyLysThrGlyHisCysGluArgGlyCysProGlnAspArgPheGly 70

236 AAA......CAGCAC.................................... 228
        |||     ::|||
 71 LysGlyCysGluHisLysCysAlaCysArgAsnGlyGlyLeuCysHisA1 87

227 .ACGAATGGCTCCGATAAC........................GGTAAC 210
     |||||||||||| |||                         |||
 87 aThrAsnGlySerCysProLeuGlyTrpMetGlyProHisCysG 104

209 ............ATCGACGCCATTGGCGTGTTAGTCTTGTACAAACT 174
                ||||||||||||||||||| ||:::::::
104 luHisAlaCysProAlaGlyTyrGlyAlaCysLeuLeuGluCys 120

173 ....................ACGGCTTCACCGACGCTCCGAGGGGCATGTTT 142
                        :::::::||||||||||  ||||||||||
121 SerCysGlnAsnAsnGlySerCysSerProThrSer...GlyAlaCysLe 136

141 GAGAGACCCTCA........GCTTGGAGGTGTCTTTGTCCT 129
        | |||::::||           |||  ||||||||||||||||
136 uCysGlyProGlyPheTyrGlnAlaCysGluAspThrCysProAlaG 153

128 ....AAAGGGCAGGGACTGTTTCCATTATGCCGTGCCAGCTACAGTAG 84
        ||:::||||             :::||||||
153 lyPheHisGlySerGlyCysGlnArgValCysGlnGlnGlyAla 169

 83 CCGTGTCCG......GCTTGGAGGTGTCTTTGTCCT 54
        |||||       :::   ||||||||||||
170 ProCysAspProValSerGlyArgCysLeuCysPro 181
```

seq_name: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:AAY76077

seq_documentation_block:
ID   AAY76077 standard; Protein; 299 AA.
XX
AC   AAY76077;

```
XX
DT   27-MAR-2000 (first entry)
XX
DE   Murine EGF family homologue, SEQ ID NO:332.
XX
KW   Skin; dermal papilla; keratinocyte; neonatal foreskin fibroblast;
KW   embryonic skin cell; keratinocyte stem cell; transit amplifying cell;
KW   secreted; transmembrane; inflammation; cancer; neurological disease;
KW   angiogenesis; tumour vascularisation; growth disorder;
KW   developmental disorder; skin wound; hair follicle disorder;
KW   anti-inflammatory; cytostatic; neuroprotective; vulnery.
XX
OS   Mus sp.
XX
PN   WO955865-Al.
XX
PD   04-NOV-1999.
XX
PF   29-APR-1999; 99WO-NZ00051.
XX
PR   29-APR-1998; 98US-0069726.
PR   09-NOV-1998; 98US-0188930.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
XX
PI   Strachan L, Sleeman M, Watson JD, Onrust R, Kumble A, Murison JG;
XX
DR   WPI; 2000-072177/06.
XX
PT   Novel polynucleotides useful for the treatment of various conditions
PT   including wounds and cancer -
XX
PS   Claim 4; Page 194; 235pp; English.
XX
CC   The invention relates to novel nucleic acid sequences derived from rat
CC   dermal papilla, human keratinocytes and neonatal foreskin fibroblasts,
CC   and mouse embryonic skin, keratinocyte stem cells and transit amplifying
CC   cells. Polypeptides of the invention may be used to treat inflammation,
CC   cancer and neurological diseases. The proteins may be used to stimulate
CC   the growth and motility of keratinocytes, to inhibit the growth of
CC   cancer cells, to modulate angiogenesis and tumour vascularisation, to
CC   modulate skin inflammation, to modulate epithelial cell growth and to
CC   inhibit binding of HIV-1 to leukocytes. The invention may also be used
CC   to treat growth and developmental defects, skin wounds and hair follicle
CC   disorders. Sequences AAY75942-Y76123 represent polypeptides encoded
CC   by cDNA sequences derived from several mouse, rat or human skin cell
CC   types. Sequences AAY75942-Y75947, AAY76020-Y76021, AAY76094-Y76104 and
CC   AAY76119 are proteins with an N-terminal signal sequence, indicating
CC   that they are secreted. Sequences AAY75986-Y75989, AAY76061-Y76071,
CC   AAY76106-Y76109 and AAY76121-Y76122 are proteins with one or more
CC   putative transmembrane domains.
XX
SQ   Sequence 299 AA;

alignment_scores:
     Quality:    68.50              Length:    162
     Ratio:       1.123             Gaps:        8
Percent Similarity: 37.654    Percent Identity: 24.074

alignment_block:
US-09-540-234-1/rev x AAY76077   ..

Align seg 1/1 to: AAY76077   from: 1   to: 299

371 TGTCCCCAGACGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA 322
      ||||||||| |||            |||
 21 CysProGlnGlyArgPheGlyLeuProSerCysAlaHisValCysThrCysGl 37

321 GGAAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG 272
    ::: ||||||||| |||||||||| |||::::::||| :|:::::||| |
 37 yGlnGlyAlaAlaCysAspProValSerGlyThrCysIleCysProProG 54
```

```
271 GA.....................GTACAGCAAGGAAGTCATCAAGCGCTACACAGT 237
    ||                      ::::::|||         |||       :::
 54 lyLysThrGlyGlyHisCysGlyCysProGlnArgGlyCysProGlnAspArgPheGly 70

236 AAA........CAGCAC.. 228
    |||        ::|||
 71 LysGlyCysGluHisLysCysAlaCysArgAsnGlyGlyLeuCysHisAl 87

227 .ACGAATGGCTCCGATAAC.. 210
    ||||||||||| ::
 87 aThrAsnGlySerCysSerCysProLeuGlyTrpMetGlyProHisCysG 104

209 ...........ATCGACGCCATTGGCGTTGTTAGTCTTGTACAAACT 174
                     |||::::::|    |||::::::
104 luHisAlaCysProAlaGlyArgTyrGlyAlaAlaCysLeuLeuGluCys 120

173 ......ACGGCTTCACCGACGTCTCGAGGGGCATGTTT 142
            :::::||||||||||    ||||||||||
121 SerCysGlnAsnAsnGlySerCysGluProThrSer...GlyAlaCysLe 136

141 GAGACGACCCTCA.. 129
    |  ||:::
136 uCysGlyProGlyPheTyrGlyGlnAlaCysGluAspThrCysProAlaG 153

128 ....AAAGGGCAGGGACTGTTTTCCATTATGCCGTGCCAGCTACAGTAG 84
        ||:::|||         :::          ||||||
153 lyPheHisGlySerGlyCysGlnArgValCysGlnArgGlnGlnGlyAla 169

 83 CCGTGTCCG......GCTTGGAGGTGTCTTTGTCCT 54
    ||||||       ||||||||||||||||
170 ProCysAspProValSerGlyArgCysArgCysLeuCysPro 181
```

THIS PAGE BLANK (USPTO)

OM of: US-09-540-234-1 to: SPTREMBL_16:*   out_format : pfs

Date: Sep 7, 2001   5:10 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09540234/runat_07092001_145855_19463/app_query.fasta_1.507
 -DB=SPTREMBL_16 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=0.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -MINLEN=0 -MAXLEN=2000000000
 -USER=US09540234_@CGN1_1_85 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-540-234-1
Query length: 444
Database: SPTREMBL_16:*
Database sequences: 425026
Database length: 132305027
Search time (sec): 116.920000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	! Documentation
sp_organelle:Q9MR87	+	90.00	161.89	0.2463	353	! Q9mr87 melosira ambigua. cytoch
sp_organelle:Q9TE67	+	87.00	156.00	0.5248	353	! Q9te67 rhizosolenia setigera. c
sp_bacteria:O83713	-	85.50	156.18	0.7662	236	! O83713 treponema pallidum. cons
sp_invertebrate:O16717	+	85.50	147.84	0.7655	689	! O16717 caenorhabditis elegans
sp_invertebrate:Q9V7I4	-	84.50	140.94	0.9845	1298	! Q9v7i4 drosophila melanogast
sp_invertebrate:Q9NGV2	-	84.50	140.57	0.9844	1361	! Q9ngv2 drosophila melanogast
sp_organelle:Q9TE70	+	84.00	150.10	1.12	353	! Q9te70 ditylum brightwellii. cy
sp_organelle:O47466	+	82.00	146.17	1.85	353	! O47466 eustigmatos magnus. cyto
sp_organelle:O47471	+	82.00	146.17	1.85	353	! O47471 ophiocytium majus. cytoc
sp_organelle:O21352	+	79.00	142.22	3.94	275	! O21352 euhadra herklotsi. cytoc
sp_organelle:O99384	+	79.00	140.27	3.94	353	! O99384 colpomenia bullosa. cyto
sp_organelle:Q9TE69	-	78.50	139.29	4.47	353	! Q9te69 fragillaria striatula. cy
sp_invertebrate:Q9NC83	+	78.00	144.04	5.08	169	! Q9nc83 strongylocentrotus pur
sp_organelle:O99381	+	78.00	138.31	5.07	353	! O99381 pseudochorda nagaii. cyt
sp_organelle:O99382	+	78.00	138.31	5.07	353	! O99382 undaria pinnatifida. cyt
sp_organelle:Q9MJ87	+	78.00	137.22	5.07	406	! Q9mj87 terebratulina pacifica.
sp_organelle:Q9T9P4	+	78.00	135.39	5.07	513	! Q9t9p4 terebratulina retusa. cy
sp_organelle:O21002	+	78.00	135.36	5.07	515	! O21002 branchiostoma lanceolatu
sp_plant:Q96397	+	78.00	133.67	5.07	640	! Q96397 chlamydomonas reinhardti
sp_invertebrate:O45936	+	77.50	133.10	5.75	607	! O45936 caenorhabditis elegans
sp_virus:O90368	+	77.50	122.03	5.75	2513	! O90368 o'nyong-nyong virus. nc
sp_organelle:O99385	+	77.00	136.34	6.53	353	! O99385 tribonema marinum. cytoc
sp_organelle:O47467	+	77.00	136.34	6.53	353	! O47467 heterococcus caespitosus
sp_plant:Q9LJ36	-	77.00	134.04	6.53	474	! Q91j36 oryza sativa (rice). hyp
sp_organelle:P92800	+	77.00	133.20	6.53	528	! P92800 pylaiella littoralis. cy
sp_organelle:Q9TE66	+	76.50	135.36	7.41	353	! Q9te66 skeletonema costatum. cy
sp_organelle:Q9TE64	.	76.50	135.36	7.41	353	! Q9te64 thalassiosira nordenskio
sp_organelle:Q9G407	+	76.50	132.48	7.40	511	! Q9g407 lithobius forficatus. cy
sp_plant:Q9MQ04	+	76.00	136.61	8.40	265	! Q9mq04 tetraselmis aff. maculat
sp_invertebrate:O16955	-	76.00	134.46	8.40	349	! O16955 caenorhabditis elegans
sp_human:Q9P234	+	76.00	129.76	8.40	638	! Q9p234 homo sapiens (human). k
sp_human:Q9HAM2	-	76.00	128.91	8.40	712	! Q9ham2 homo sapiens (human). cd
sp_organelle:P92619	+	75.50	136.56	9.53	235	! P92619 chlorella vulgaris. cyto
sp_organelle:Q9TE71	+	75.50	133.39	9.53	353	! Q9te71 cylindrotheca closterium
sp_plant:O23578	+	75.50	132.76	9.53	383	! O23578 arabidopsis thaliana (mo
sp_organelle:O63273	+	75.00	133.08	10.81	324	! O63273 cataglyphis velox. cytoc
sp_organelle:Q9MJ93	+	75.00	131.32	10.81	406	! Q9mj93 jolonica nipponica. cytc
sp_organelle:Q9XP16	+	75.00	131.24	10.81	410	! Q9xp16 microplitis maturus. cyt
sp_organelle:Q9G4C3	+	75.00	129.65	10.81	503	! Q9g4c3 thraustochytrium aureum.
sp_invertebrate:O17047	-	74.50	126.02	12.25	707	! O17047 caenorhabditis elegans

sp_organelle:O99383	+	74.00	130.44	13.91	353	! O99383 ectocarpus sp. cytoch
sp_organelle:Q9TE65	+	74.00	130.44	13.91	353	! Q9te65 thalassionema nitzsch
sp_organelle:Q9MJ88	+	74.00	129.35	13.91	406	! Q9mj88 terebratulina crossei
sp_bacteria:Q9PC48	+	74.00	128.60	13.91	447	! Q9pc48 xylella fastidiosa. P
sp_organelle:Q9G870	+	74.00	127.97	13.91	485	! Q9g870 malawimonas jakobifor

seq_name: sp_organelle:Q9MR87

seq_documentation_block:
ID Q9MR87   PRELIMINARY;   PRT;   353 AA.
AC Q9MR87;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Melosira ambigua.
OG Mitochondrion.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Coscinodiscophycidae; Melosirales; Melosiraceae; Melosira.
OX NCBI_TaxID=70449;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIES20;
RX MEDLINE=98440755; PubMed=9767683;
RA Inagaki Y., Ehara M., Watanabe K.I., Hayashi-Ishimaru Y., Ohama T.;
RT "Directionally evolving genetic code: the UGA codon from stop to
RT tryptophan in mitochondria.";
RL J. Mol. Evol. 47:378-384(1998).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AB009418; BAA96357.1; -.
DR InterPro; IPR000883; -.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASEI.
DR PROSITE; PS00077; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW Respiratory chain; Transmembrane.
FT NON_TER   1   1
FT NON_TER   353   353
SQ SEQUENCE 353 AA; 38170 MW; E2521D002FDCB75B CRC64;

alignment_scores:
          Quality: 90.00   Length: 161
          Ratio: 1.139   Gaps: 8
Percent Similarity: 49.068   Percent Identity: 25.466

alignment_block:
US-09-540-234-1 x Q9MR87

Align seg 1/1 to: Q9MR87 from: 1 to: 353

107 TGGAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTCAAACATG........ 148
                 |||:||||||| :::::||||| ::::|||:::||:::::::
 90 TrpLeuProSerLeuLeuLeuLeuIleSerSerValLeuAlaGl 106

149 ....CCCCTCGAGACG..... 160
                                             |||||||::::|||
106 uAlaGlyValGlyThrGlyTrpThrValTyrProLeuSerThrValS 123

161 ....................................TCGGTGAAGCCGTAGTTGTACAAGACTAACAA 193

```
FT   NON_TER          1      1
FT   NON_TER        353    353
SQ   SEQUENCE       353 AA;   38138 MW;   65D9C11AA4D48870 CRC64;

alignment_scores:
                    Quality:   87.00                  Length:      161
                      Ratio:    1.160                   Gaps:        8
        Percent Similarity:  46.584     Percent Identity:  27.329

alignment_block:
US-09-540-234-1 x Q9TE67

Align seg 1/1  to: Q9TE67   from:  1   to: 353

107  TGGAAACAGTCCCTGCCCTTTGAGGTGTCGTCTCAAACATGCCCCTCGA  156
     |||    || ||:|:||||  :::||:|||| |    ::|||: :::|||||  ||
90   TrpLeuLeuProSerLeuLeuLeuLeuValAlaSerMetLeuAlaGl   106

157  GACG...............TCGGTGAAGCCGTAGTTTGTACAAGACT  188
     ||||               :::|||| ::| | |:|::|||:||::||:||
106  uAlaGlyValGlyThrGlyTrpThrValTyrPro.ProLeuSerSerGly  122

189  AACAACGCCAATGGCGTCGATGTT......................  212
     || |:|||  ::|:|||:::|||  |||
123  AsnSerHisSerGlyAlaSerValAspLeuAlaIlePheSerLeuHisLe  139

213  ...........................ATCGGAGCCATTCGTGCGTGTTTACTGTGT  243
                               ::||||||||||| :|: |||||||:|:||
139  uSerGlyAlaAlaSerIleLeuGlyLeuGlyAlaIleAsnPheIleCysThrIleL  156

244  AGCGGCTTGATGACTTCCTTGCTGTACTCC.................GCT  275
     ::::|||:|||||||||||   |||:|||||||                 ||
156  euAsnMetArgValLysGlyLeuPheMetHisArgLeuProLeuPheVal  172

276  TGGAGCCATTGGAGTGCTGTTTACCGTGCCAGCACTGCGAAGACTTCCTTGC  325
     |||||||| |||  :::|||||    ||||||:||:|||||:::::||
173  TrpSerIleLeuIleThrAlaValLeuLeuLeuSer.....LeuPr  187

326  TGTACTCCTGGCAGCATAGTCTGTTTGCTG................  356
     |||||||||||:|||::||||:|| ::||||
187  oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnT  204

357  .....................CCTGGTCTGGGGACACCTGTACTACCAGTACT  389
                         ||| |||  || ||||||:||:|||||||||
204  hrThrPhePheAspProAlaGlyGlyGlyAspProValLeuTyrGlnHis  220

390  TCGTTT................CCCTTTGTAGTGGTGTACTACTTCTT......  419
     ||:||                |||  ||:   |||  |:|||:::::|||
221  LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGl  237

420  .............AGCCATGTACTCGTGTCT  437
                  |||||:|||||:::|||||
237  yPheGlyIleIleSerHisIleIleValSer  247

seq_name: sp_bacteria:O83713

seq_documentation_block:
ID   O83713         PRELIMINARY;         PRT;        236 AA.
AC   O83713;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   TP0731.
OS   Treponema pallidum.
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS.
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
```

seq_name: sp_organelle:Q9TE67

seq_documentation_block:
ID   Q9TE67   PRELIMINARY;   PRT;   353 AA.
AC   Q9TE67;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Rhizosolenia setigera.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC   Rhizosoleniophycidae; Rhizosoleniales; Rhizosoleniaceae; Rhizosolenia.
OX   NCBI_TaxID=3005;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCMP 1330;
RA   Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT   "Phylogenetic analysis of diatom coxI genes and implications of
RT   fluctuating GC content on mitochondrial genetic code evolution.";
RL   Curr. Genet. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB020226; BAA86611.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.

123 erSerHisSerGlyProAlaVal.AspLeuAlaIlePheSerLeuHisLe 139
     :::|||| ::: ::::::|::::::| :::
194 CGCCAATGGCGTCGATGTTATCCGAGCCATTCGTGCGTGTTTACTGTGT 243
     :::::::::::::|||||||||||||:: ||||||

139 uSerGlyAlaAlaSerIleLeuGlyLeuGlyAlaIleAsnPheIleCysThrIleL 156
     :||||||||||||::::::::||||||||| : ||||:::
244 AGCGGCTTGATGACTTCCTTGCTGTACTCC.................GCT 275

156 euAsnMetArgValLysGlyLeuPheMetHisLysLeuProLeuValVal 172
     :::||||||||||||||::|||||::||||||||:||
276 TGGAGCCATTGGAGTGCTGTTTACCGTGCCAGCACTGCGAAGACTTCCTTGC 325

173 TrpSerIleLeuIleThrAlaValLeuLeuLeuSer.....LeuPr 187
     ::::::|||| |||:::
326 TGTACTCCTGGCAGCATAGTCTGTTTGCTG................ 356

187 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnT 204
     |||||| ||||||::::|:|||::||
357 .....................CCTGGTCTGGGGACACCTGTACTACCAGTACT 389

204 hrThrPhePheAspProAlaGlyGlyGlyAspProValLeuTyrGlnHis 220
     |||||| :|| ||||||::::|||||
390 TCGTTT................CCCTTTGTAGTGGTGTACTACTTCTT...... 419

221 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProAl 237
     |||||| :|||::|||::|||||
420 .............AGCCATGTACTCGTGTCT 437

237 aPheGlyIleIleSerHisIleIleValSer 247

RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
DR   EMBL; AE001245; AAC65700.1; -.
DR   TIGR; TP0731; -.
DR   InterPro; IPR000086; -.
DR   Pfam; PF00293; mutT; 1.
DR   PRINTS; PR005502; NUDIXFAMILY.
DR   PROSITE; PS00893; MUTT; 1.
SQ   SEQUENCE   236 AA; 27119 MW; 86B2BDD568B45EE4 CRC64;

alignment_scores:
        Quality: 85.50   Length: 128
        Ratio: 1.257   Gaps: 8
Percent Similarity: 53.125   Percent Identity: 30.469

alignment_block:
US-09-540-234-1/rev x 083713

Align seg 1/1 to: 083713 from: 1 to: 236

392 CGAAGTACTGTA.....GTACACAGGTGTCCCCAGACCAGGCAGCAAAC 349
     |||::||||||| :|||||||| :|| :: :::|||:|||
 23 ArgGluThrValGlnProAlaHisArgAlaLysGluGluLysGlnAsn 39

348 AGACTATGCTGCCAGCGAGTACAGCCAAGGAAGT..........CTTC 312
     :|| ||| ||||::|| |||
 39 pAlaHisLeuAlaTrpGluValArgLysArgHisGluAlaCysArgLeuA 56

311 GAGTGGCTGCACGGTAAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCA 262
     ||||::||| :|||:| :::::| ||| |||:|:||||:|||:::
 56 rgValPheHisValGlnGluLeuGluSerValSerProArgLysThrVal 72

261 AGGAAGTCATCAAGCGCTACACAGTAAACAGCACAGAATGGCTC..... 217
     |||::|| :: ||||| |||||:||:|:||
 73 ArgArgPheValThrLeu........ThrAlaProGluTrpValIleVa 86

216 ....CGATACATCGA..............CGCC 201
     ||| ::||| |||
 86 lValProHisValMetGluArgAlaGlnArgPhePheValMetValArgG 103

200 ATTGGCGGTTGTTAGTCT........TGTACAAACTAC......GGCTTC 166
     ::||||||||| ||| ||||||:|::::| |||
103 lnTrpArgCysGlySerGlnThrValCysThrGluPheProGlyGlyVal 119

165 ACCGACGTCTCG.............AGGGGCATGTTTGA 140
     |||::||||| ||::|||||
120 IleAspAlaGlyHisProGluAlaAlaAlaAlaArgArgGluLeuPheGl 136

139 GACGACCCTCAAAAGGGCAGGGACTGTTTTCCAT 106
     | ||| ::||||||:::||||||| |||
136 uGluThrGlyArgArgAlaSerSerLeuAlaHis 147

OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier H., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Graves T., Wohldmann P., Clarke K.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016681; AAB86178.1; -.
DR   InterPro; IPR000168; -.
DR   InterPro; IPR003003; -.
DR   Pfam; PF01604; 7tm_5; 2.
SQ   SEQUENCE   689 AA; 78022 MW; 6880DA7F47557A23 CRC64;

alignment_scores:
        Quality: 85.50   Length: 121
        Ratio: 1.276   Gaps: 5
Percent Similarity: 55.372   Percent Identity: 27.273

alignment_block:
US-09-540-234-1 x O16717

Align seg 1/1 to: O16717 from: 1 to: 689

114 CAGTCCCTGCCC...TTTTGAGGTCGTCTCAAACATGCCCTCGAGACG 160
     :::::|||||| ::||| ::||| ||||:: |||
134 GluAsnLeuProGluTyrLeuSerSerPheGlnSerLeuPro....... 147

161 TCGGTGAAGCCGTAGTTGTACAAGACTAACAACGCCAATGGCGTGCGATG 210
     |||||||||||||||||:|| ::|||||||||:: ||
148 ......AsnPheSerIleTyrGluAlaAsnSerLeuIlePheIleValV 162

211 TTATCGGAGCCATTCGTCGTCGTGTTTACTGTGTAGCCGTTGATGACTTCC 260
     ||||| |||:: ||: ::::|||:::||||::
162 aIIleThrAlaValThrGlyAlaPheIleAlaPheThrPhePheThrAla 178

261 TTGCTGTACTCCGCTGGAGCCATTGAGTGCTGTT.............. 296
     :::||||||| ::: ||| :::
179 IleLeuTyrAsnIlePheArgMetLeuGlyPheMetLysLeuGlnMetSe 195

297 ...................TACCGTGCAGCACTCGAAGACTTCCTTGCTG 327
     :::|||||||||||| :::|||
195 rAlaThrThrTyrLysGlnHisArgAlaAlaValTrpSerLeuIleAlaG 212

328 TACTCCTGGCAGCATAGTCGTTTGCGCCTGGTCTCGGGGACACCTGTG 377
     :::||| |||||||| ||::::|||| |||:::
212 lnPheAlaThrSerIleIleCysPheSerPro..........ProIle 224

378 TACTACAGTACTCGTTCCCTTTGTAGTGGTACTACTTCTTAGCCATGT 427
     ::: |||||||| :: ::| || ||
225 Phe..........LeuValPheValValPheValPheGlnLeuProHisAl 237

seq_name: sp_invertebrate:O16717

seq_documentation_block:
ID   O16717   PRELIMINARY;   PRT;   689 AA.
AC   O16717;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   F22E5.15 PROTEIN.
GN   F22E5.15.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.

```
428 ACTCGTCTCGTT 440
    :  ||| |||
237 aGlnValIleVal 241
```

seq_name: sp_invertebrate:Q9V7I4

alignment_scores:
```
          Quality:   84.50          Length:    109
            Ratio:    1.482            Gaps:      5
Percent Similarity:  52.294  Percent Identity:  28.440
```

alignment_block:
US-09-540-234-1/rev x Q9V7I4

Align seg 1/1 to: Q9V7I4 from: 1 to: 1298

```
365 CAGACCAGGCAGCAAACAGACTATGCTGCCACGGAGTACAGCAAGGAAGT 316
    ||||||  |||||||| :::: ::::::  :::  :::::::
239 GlnThrTyrGlnGlnThrSerAsnThrAlaLysAsnArgArgLy 255

315 CTTCGAGTGCTGCACGGTAAACACCACTCAAATGCTCCAAGCCGGAGTAC 266
    ::  ||| |||:::::!|||||
255 sTyrProAsnValThrSerAsnLysValGlnMet............... 266

265 AGCAAGGAAGTCATCAAGCCGCTACACAGTAAACAGCCACGAATGGCTCC 216
                           |||  :::!||||||
267 ...................HisIleThrArgGluThrAsnArgLeu 275

215 GATAACATCGACGCCATT....GGCGTTGTTAGTCTTGTACAAACTAC 172
    ::||||||:::|||                       |||||||
276 GluAsnIleSerAlaGluProAlaValValGlyAlaAlaMetThrTh 292

171 G......GCTTCACCGACGTCTCGAGGGCATGTTGAGACGACCCTCAA 128
    |    :::|||||||    ::                 |||
292 rAspArgSerThrProThrCysAsnLeuAspCys............G 304

127 AAGGGCAGGGACTGTGTTTCCATTATGCCGTGCCAGCTACAGTAGCCGTGT 78
                    |||  :::   |||       ::::
304 lySerAspGlyIle...............CysAlaLeuGluAlaThrAla 315

77 CCGGCTTGGAGGTGTCTTTGTCCTTTC 51
   :::  |||||||||||||||||
316 AlaSerSerArgCysLeuCysProPhe 324
```

seq_name: sp_invertebrate:Q9NGV2

seq_documentation_block:
```
ID  Q9NGV2         PRELIMINARY;       PRT;    1361 AA.
AC  Q9NGV2;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  SP2353.
GN  SP2353 OR CG8403.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=ENDOPLASMIC RETICULUM;
RA  Serano T.L., Pendleton J.D., Rubin G.M.;
RT  "A reverse genetic screen for genes involved in Drosophila
RT  development.";
RL  Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF239610; AAF63502.1; -.
DR  FlyBase; FBgn0034070; SP2353.
DR  InterPro; IPR000561; -.
DR  InterPro; IPR001791; -.
DR  Pfam; PF00008; EGF; 2.
DR  Pfam; PF00054; laminin_G; 3.
DR  PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR  PROSITE; PS01186; EGF_2; 2.
DR  SMART; SM00181; EGF; 1.
KW  EGF-like domain; Glycoprotein.
SQ  SEQUENCE 1361 AA;  150355 MW;  206D9F9BF9EDFE47 CRC64;
```

seq_name: sp_invertebrate:Q9V7I4

seq_documentation_block:
```
ID  Q9V7I4         PRELIMINARY;       PRT;    1298 AA.
AC  Q9V7I4;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  CG8403 PROTEIN.
GN  SP2353 OR CG8403.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA  Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
DR  EMBL; AE003808; AAF58071.1; -.
DR  FlyBase; FBgn0034070; SP2353.
DR  InterPro; IPR000561; -.
DR  InterPro; IPR001791; -.
DR  Pfam; PF00008; EGF; 3.
DR  Pfam; PF00054; laminin_G; 2.
DR  PROSITE; PS00022; EGF_1; 2.
DR  PROSITE; PS01186; EGF_2; 2.
DR  SMART; SM00181; EGF; 1.
KW  EGF-like domain; Glycoprotein.
SQ  SEQUENCE 1298 AA;  143468 MW;  587B8330C1353048 CRC64;
```

```
alignment_scores:
    Quality:    84.50          Length:    109
    Ratio:      1.482          Gaps:      5
Percent Similarity: 52.294    Percent Identity: 28.440

alignment_block:
US-09-540-234-1/rev x Q9NGV2  .

Align seg 1/1  to: Q9NGV2  from: 1  to: 1361

365 CAGACCAGGCAGCAAACAGCATGCTGCCAGCGAGTACAGCAAGGAAGT 316
    ||||| ||||||||||::  ::  :: ::::  :::::::::
302 GlnThrTyrGlnThrSerAsnThrAsnThrAlaLysAsnArgArgLy 318

315 CTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTAC 266
    ::  ||||||||| ||| ||:::|||||
318 sTyrProAsnValThrSerAsnLysValGlnMet............... 329

265 AGCAAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCC 216
                   |||       :::|||||
330 ..............HisIleThrArgGluThrAsnArgLeu 338

215 GATAACATCGACGCCATT......GGCGTTGTTAGTCTTGTACAAACTAC 172
    ::|||||||::||          ::||||||:::  :::::|||||
339 GluAsnIleSerAlaAlaGluProAlaValValGlyAlaMetThrTh 355

171 G.......GCTTCACCGACGTCTCCGAGGGGCATGTTGAGACGACCCCTCAA` 128
    :::|:||||||      |||
355 rAspArgSerThrProThrCysAsnLeuAspCys..............G 367

127 AAGGGCAGGGACTGTTTTCCATTATGCCGTGCCAGTACAGTAGCCGTGT 78
    :::::|||:||     ||| |||:::  ||||||||||||
367 lySerAspGlyIle.............CysAlaLeuGluAlaThrAla 378

77 CCGGCTTGGAGGTGTCTTTGTCCTTTC 51
   ::: |||||||||||||||||||||
379 AlaSerSerArgCysLeuCysProPhe 387

seq_name: sp_organelle:Q9TE70

seq_documentation_block:
ID   Q9TE70        PRELIMINARY;      PRT;     353 AA.
AC   Q9TE70;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Ditylum brightwellii.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC   Lithodesmiaceae; Ditylum.
OX   NCBI_TaxID=49249;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP1022-2;
RA   Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT   "Phylogenetic analysis of diatom coxI genes and implications of
RT   fluctuating GC content on mitochondrial genetic code evolution.";
RL   Curr. Genet. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB020223; BAA86608.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COXI; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1    1
FT   NON_TER     353  353
SQ   SEQUENCE   353 AA;  38145 MW;  CC5018EC930C9685 CRC64;

alignment_scores:
    Quality:    84.00          Length:    145
    Ratio:      1.105          Gaps:      8
Percent Similarity: 52.414    Percent Identity: 28.276

alignment_block:
US-09-540-234-1 x Q9TE70   ..

Align seg 1/1  to: Q9TE70  from: 1  to: 353

92 GCTGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
   ||||||||||| ||| ::: ||| ||| ||:::
109 AlaGlyThrGly...TrpThrValTyrPro.ProLeu......SerSerA 122

142 AAACATGCCCCTGCGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    ||::: |||        ::: :::||||||
122 laThrAlaHisSerGlyGlyAlaValAspLeuAlaIlePheSerLeuHis 138

192 AACGGCCAATGCGTCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGT 241
    :::::::|| ||||    ||||| :::::|| :::|:::  :|||:||
139 LeuSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleCysThrIl 155

242 GTAGGCGCTTGATGACT..................TCCTTGCTGTACTCCG 273
    :::::|:||:||          ::||
155 ePheAsnMetArgValLysSerLeuSerPheHisAsnLeuProLeuPheV 172

274 CTTGGAGCCATTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTT 323
    ::||||| |||  :::::::::::    |||::: |||:::
172 alTrpSerValLeuIleThrAlaPheLeuLeuLeuLeuSer......Leu 186

324 GCTGTACTCGCTGGCAGCATAGTCGTGTTTGCTG.............. 356
    |||||||||||::|||::::||:|||     ||||||
187 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 203

357 ..................CCTGGTCGGGACAACCTGTGTACTACAGTA 387
                      ||| ||| |||||||||||||| ||:::::
203 nThrThrPheAspProAlaGlyGlyAspProValLeuTyrGlnH 220

388 CTTCGTTT.............CCCTTGTAGTGGTACTACTTCTT... 419
    ||| |||          ::|||||:::||||||:::||:::||
220 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 236

420 .............AGCCATGTACTCGTGTCT 437
                 |||||||||||||||||||
237 GlyPheGlyIleValSerHisIleValValSer 247

seq_name: sp_organelle:O47466

seq_documentation_block:
ID   O47466        PRELIMINARY;      PRT;     353 AA.
AC   O47466;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Eustigmatos magnus.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Eustigmatophyceae; Eustigmatos.
OX   NCBI_TaxID=73014;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 860/2;
RX   MEDLINE=97383241; PubMed=9236270;
RA   Ehara M., Hayashi-Ishimaru Y., Inagaki Y., Ohama T.;
RT   "Use of a deviant mitochondrial genetic code in yellow-green algae as
RT   a landmark for segregating members within the phylum.";
RL   J. Mol. Evol. 45:119-124(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 860/2;
RX   MEDLINE=97451042; PubMed=9302324;
RA   Inagaki Y., Hayashi-Ishimaru Y., Ehara M., Igarashi I., Ohama T.;
RT   "Algae or protozoa: phylogenetic position of euglenophytes and
RT   dinoflagellates as inferred from mitochondrial sequences.";
RL   J. Mol. Evol. 45:295-300(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 860/2;
RX   MEDLINE=98004465; PubMed=9342410;
RA   Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT   "A deviant mitochondrial genetic code in prymnesiophytes (yellow-
RT   algae): UGA codon for tryptophan.";
RL   Curr. Genet. 32:296-299(1997).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB000205; BAA24969.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     353    353
SQ   SEQUENCE   353 AA;   37913 MW;   FF491AFB46532F00 CRC64;
```

alignment_scores:
```
        Quality:    82.00         Length:     146
          Ratio:     1.079          Gaps:       8
Percent Similarity:  52.055   Percent Identity:  28.767
```

alignment_block:
US-09-540-234-1 x O47466   ..

Align seg 1/1 to: O47466 from: 1 to: 353

```
 92 GCTGGCACGGCATAATGGAAAACAGTCCCCGCCCTTTGAGGGTCGTCTC 141
    ||||||||||| ||  ::  ||| || ||| || |||  :::|| |||||
109 AlaGlyThrGly...TrpThrValTyrProLeuSerValGlnAl 124

142 AAACATGCCCCTGCAGAGCGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    : ::::        |||||  ::: :::::::::::|| :: ||:
124 aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGCTGTTTACTGT 241
    ||| |||||||||| :||||||||||||:::  ::: :||||| |||||
139 LeuAlaGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrVa 155

242 GTAGGCGTTG.........ATGACTTCCTTGCTGTACTCCG 273
    ||| :  :::::          |||           |||
155 lPheAsnMetArgGlyGlyIleThrMetHisArgProLeuPheV 172
```

```
274 CTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    :::|||::: |||           :::|||:::::::::::::    |||
172 aTrpAlaValLeuIleThrAlaPheLeuLeuValLeuSer......Leu 186

324 GCTGTACTCGCTGCCGCAGCATAGTCTGTTTGCTG................ 356
    ||||||||:||||::||||||:::||:::     |||||
187 ProValLeuAlaGlyAlaIleThrMetLeuThrAspArgAsnPheAs 203

357 ................CCTGCTCGGGACACCTGGTACTACAGTA 387
                    ||| ||| |||||| ||:::
203 nThrThrPheAspProAlaGlyGlyAspProValLeuTyrGlnH 220

388 CTTCGTTT..........CCCTTTGTAGTGGTACTACTTCTT... 419
    ||:                ::||||||||||:::|||
220 isLeuPheTrpPhePheGlyHisProGluvalTyrIleLeuIleLeuPro 236

420 ............AGCCATGTACTCGTCGTCTGTT 440
                ||||||||:: ||| :::
237 AlaPheGlyIleIleSerHisValValSerSerLeu 248
```

seq_name: sp_organelle:O47471

```
seq_documentation_block:
ID   O47471        PRELIMINARY;      PRT;    353 AA.
AC   O47471;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Ophiocytium majus.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Xanthophyceae; Ophiocytium.
OX   NCBI_TaxID=73019;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 855/1;
RX   MEDLINE=97383241; PubMed=9236270;
RA   Ehara M., Hayashi-Ishimaru Y., Inagaki Y., Ohama T.;
RT   "Use of a deviant mitochondrial genetic code in yellow-green algae as
RT   a landmark for segregating members within the phylum.";
RL   J. Mol. Evol. 45:119-124(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 855/1;
RX   MEDLINE=97451042; PubMed=9302324;
RA   Inagaki Y., Hayashi-Ishimaru Y., Ehara M., Igarashi I., Ohama T.;
RT   "Algae or protozoa: phylogenetic position of euglenophytes and
RT   dinoflagellates as inferred from mitochondrial sequences.";
RL   J. Mol. Evol. 45:295-300(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 855/1;
RX   MEDLINE=98004465; PubMed=9342410:
RA   Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT   "A deviant mitochondrial genetic code in prymnesiophytes (yellow-
RT   algae); UGA codon for tryptophan.";
RL   Curr. Genet. 32:296-299(1997);
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB000210; BAA24974.1; -.
```

DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme: Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER   1   1
FT   NON_TER   353   353
SQ   SEQUENCE   353 AA;   37913 MW;   5C7FDDE1F3AE6694 CRC64;

alignment_scores:
            Quality: 82.00   Length: 146
              Ratio: 1.079   Gaps: 8
Percent Similarity: 52.055   Percent Identity: 28.767

alignment_block:
US-09-540-234-1 x 047471

Align seg 1/1   to: 047471   from: 1   to: 353

 92 GCTGGCACGGCATAATGAGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    ||||||||||::: ||| :::||| ||| ::: || ::
109 AlaGlyThrGly...TrpThrValTyrProLeuSerSerValGlnAl 124

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAAC 191
    ::: |||||| :::   ::::::::::   :::
124 aHisSer......GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGT 241
       ||::::::::::::::||||||||::   ::: ||||||
139 LeuAlaGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrVa 155

242 GTAGGCGCTTG..............ATGACTTCCTTGCTGCTACTCCG 273
       ::::::: ||| :::||||
155 lPheAsnMetArgGlyHisMetHisArgLeuProLeuPheV 172

274 CTTGGAGCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    ::||| ||| :::::::::|| :::|||:: ||||
172 alTrpAlaValLeuIleThrAlaPheLeuLeuLeuSer......Leu 186

324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG................ 356
    |||||||||||| |||:::||||:::|||
187 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 203

357 ................CCTGCTGGGACACCTGTGTACTACAGTA 387
                        ||| |||||| |||::: |||::::
203 nThrThrPhePheAspProAlaGlyGlyAspProValLeuTyrGlnH 220

388 CTTCGTT...........CCCTTTGTAGTGGTACTACTTCTT... 419
       ||| ||| ||| ||| :::||||||||
220 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 236

420 ...........AGCCATGTACTCGTCTGTT 440
              |||||||||| :::   ::: ||
237 AlaPheGlyIleIleSerHisValValSerSerLeu 248

seq_name: sp_organelle:021352

seq_documentation_block:

CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; Z71701; CAA96383.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER   275   275
SQ   SEQUENCE   275 AA;   30368 MW;   7B104A2A1FD2D819 CRC64;

alignment_scores:
            Quality: 79.00   Length: 143
              Ratio: 1.129   Gaps: 8
Percent Similarity: 48.951   Percent Identity: 28.671

alignment_block:
US-09-540-234-1 x 021352

Align seg 1/1   to: 021352   from: 1   to: 275

 92 GCTGGCACGGCATAATGAGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    |||||||||||| ||| :::||| |||::: ||||||
118 AlaGlyThrGly...TrpThrValTyrPro.ProLeu......SerSerL 131

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAAC 191
    ||| ||| :::   :::::::::: |||:::
131 euThrGlyHisSerGlyAlaSerValAspLeuAlaIlePheSerLeuHis 147

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGT 241
    :::||||||||||||::   :::||||||||||::
148 LeuAlaGlyMetSerSerIleLeuGlyAlaIleAsnPheIleThrThrIl 164

242 GTAGGCGCTTG..................ATGACTTCCTTGCTGCTACTCCG 273
       ::::::: ||| :::||||
164 ePheAsnMetArgAlaProGlyValThrMetGluArgLeuSerLeuPheV 181

274 CTTGGAGCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    ::|||||| ||| :::::::|| :::|||:: |||
181 alTrpSerIleLeuValThrValPheLeuLeuLeuSer......Leu 195

324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG................ 356
    ||||||||||||| |||:::||||:::|||
196 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 212

357 ................CCTGCTGGGACACCTGTGTACTACAGTA 387
                        ||| |||||| |||::: |||::::
212 nThrSerPhePheAsnProAlaGlyGlyAspProIleLeuTyrGlnH 229

388 CTTCGTT...........CCCTTTGTAGTGGTACTA......... 413
       ||| ||| ||| :::|||
229 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleGlnPro 245

414 ........CTTCTTAGCCATGTACTC 431
          |||||| :::|||
246 GlyPheGlyIleIleSerHisSerIleLeu 254

```
seq_name: sp_organelle:O99384

seq_documentation_block:
ID   O99384        PRELIMINARY;        PRT;        353 AA.
AC   O99384;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Colpomenia bullosa.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;
OC   Scytosiphonaceae; Colpomenia.
OX   NCBI_TaxID=74371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MURORAN, HOKKAIDO, JAPAN;
RA   Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA   Ohama T.;
RT   "Phylogenetic analyses of heterokont algae with emphasis on the AUA
RT   codon reassignment in mitochondria.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AF037995; AAC94981.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     353    353
SQ   SEQUENCE    353 AA;   38077 MW;   2E00FFC3ABCA7286 CRC64;

alignment_scores:
              Quality:  79.00              Length:    145
              Ratio:    1.053              Gaps:        8
Percent Similarity:  51.724       Percent Identity:  27.586

alignment_block:
US-09-540-234-1 x O99384

Align seg 1/1  to: O99384  from: 1  to: 353

 92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    ::::   :::       :::    ::  ::::::::::::::::::::::::
109 AlaGlyThrGly...TrpThrValTyrProProLeuSerGlyIleGlnAl 124

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCCTAGTTTGTTACAAGACTAAC 191
     :::           ||||  :::  ::::::::::::::::::::::::::
124 aHisSer......GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138

192 AACGCCAATGGCGTCCGATGTTATCGGAGCCAATTCGTGTGCTGTTTACTGT 241
    ::::::::::::::::::::::::||||||||:::  ::: ||||||  ::: :
139 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleLeuGlyAsnPheIleThrThrIle 155
```

```
seq_name: sp_organelle:Q9TE69

seq_documentation_block:
ID   Q9TE69        PRELIMINARY;        PRT;        353 AA.
AC   Q9TE69;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Fragilaria striatula.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Bacillariophyta; Fragilariophyceae;
OC   Fragilariophycidae; Fragilariales; Fragilariaceae; Fragilaria.
OX   NCBI_TaxID=33645;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 1029-18;
RA   Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT   "Phylogenetic analysis of diatom coxI genes and implications of
RT   fluctuating GC content on mitochondrial genetic code evolution.";
RL   Curr. Genet. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB020224; BAA86609.1; -.
DR   InterPro; IPR000883;
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     353    353
SQ   SEQUENCE    353 AA;   38203 MW;   72F9296DBC214ECF CRC64;

alignment_scores:
              Quality:  78.50              Length:    130

242 GTAGCGCTTG...............................ATGACTTCCTTGCTGTACTCCG 273
    : :::::                                  |||         ::::::
155 ePheAsnMetArgAlaProGlyMetThrMetAspArgLeuProLeuPheV 172

274 CTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    :::||||||| |||   :::||||::  ||:::::|| |||||:::::||
172 alTrpSerValLeuIleThrAlaPheLeuLeuLeuLeuSer......Leu 186

324 GCTGTACTCCGCTGGCAGCATAGTCGTTTGCTG................... 356
    ||||||||||:::||:::   |||||
187 ProValLeuAlaGlyLeuGlyIleThrMetLeuLeuThrAspArgAsnPheAs 203

357 ...............CCTGCTCGGGACACCTGTACTACAGTA 387
                   |||  ||||   :::  ||||| ||| :::
203 nThrThrPheAspProAlaGlyGlyAspProValLeuTyrGlnH 220

388 CTTCGTTT...........CCCTTTGTAGTTGGTACTACTTCTT... 419
    |||                |||||||||||:::::::||
220 lsLeuPheTrpPheGlyHisProGluValTyrIleLeuIleLeuPro 236

420 ..................AGCCATGTACTCGTGTCT 437
                      ||||||||||:::::::
237 GlyPheGlyIleValSerHisValLeuAlaThr 247
```

Percent Similarity: 51.538   Percent Identity: 26.923   Gaps: 5   Ratio: 1.172

alignment_block:
US-09-540-234-1 x Q9TE69

Align seg 1/1 to: Q9TE69 from: 1 to: 353

```
 135 TCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAA 184
     |||||| ||  :::::|||    ::    :::|||||::
 120 SerSerIleThrAlaHisSerGlyGlyAlaValAspLeuAlaIlePheSe 136

 185 GACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGT 234
     ::  :::::::::::||:|||||||||  :::
 136 rLeuHisValSerGlyAlaSerGlyLeuGlyAlaIleAsnPheIleC 153

 235 TTACTGTGTAGCGCTTGATGACTTCCTTGCTG........ 266
     ||||: :::: ||||   ::::::|||   ......Cy 107
 153 ysThrIlePheAsnMetArgValLysSerLeuSerPheHisLysLeuPro 169

 267 TACTCCGCTTGGAGCCATTTGAGTGCGTGTTTACCGTGCCAGCACTCGAAGA 316
     :::||||||  |||      ||::::|||:::||
 170 LeuPheValTrpPerValLeuIleThrAlaPheLeuLeuLeuLeuSer.. 185

 317 CTTCCTTGCTGTACTCGCTCGGCAGCATAGTCTGTTTGCTG......... 356
     ||| |||||||||::|||::|| ||  ||||||
 186 ...LeuProValLeuAlaGlyAlaIleAlaIleThrMetLeuLeuThrAspArgA 201

 357 ..............................CCTGGTCTGGGACACCTGTGTAC 380
                                    ||||  ||||||
 201 snPheAsnThrThrPhePheAspProAlaGlyGlyGlyAspProValLeu 217

 381 TACAGTACTTCGTTT............CCCTTTGTAGTGGTACTA.. 413
     |||::::               ||       ||  ::||||
 218 TyrGlnHisLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuIleIl 234

 414 ..............CTTCTTAGCCATGTACTCGTGTCT 437
                   ::||||||||||:::||||||||||
 234 eValProGlyPhePheGlyIleIleSerIleIleSerHisValIleValSer 247
```

seq_name: sp_invertebrate:Q9NC83

seq_documentation_block:
```
ID Q9NC83         PRELIMINARY;    PRT;   169 AA.
AC Q9NC83;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 18.3 KDA PROTEIN.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20253941; PubMed=10793471;
RA Rast J.P., Pancer Z., Davidson E.H.;
RT "New approaches towards an understanding of deuterostome immunity.";
RL Curr. Top. Microbiol. Immunol. 248:3-16(2000).
DR EMBL; AF228878; AAF78206.1; -.
KW Hypothetical protein.
SQ SEQUENCE 169 AA; 18272 MW; D509BD9ABBE2E628 CRC64;
```

alignment_scores:
Quality: 78.00   Length: 125
Ratio: 1.322   Gaps: 5
Percent Similarity: 47.200   Percent Identity: 47.200

alignment_block:
US-09-540-234-1 x Q9NC83

Align seg 1/1 to: Q9NC83 from: 1 to: 169

```
  64 CACCTCCAAGCCGGACACGGCTACTGTAGCTGGCACGGCATAATGAAAA 113
     |||||::::|||||
  63 HisIleGlnAla.............................................. 66

 114 CAGTCCCTGCCCTTTGAGGTGCGTCTCAAACATGCCCTCGAGACGTCG 163
     ::|||  |||       :::::  ||||||||:::::|||
  67 .GlnProGlnThrThrValValArgThrAlaProArgSerThrA 83

 164 GTGAAGCCGTAGTTGTACAAGAGTAACAACGCCAATGGCGTCGATGTTA 213
     ::  :::|||    :::||  :::|||||:
  83 snAlaPheSerGlyAlaProArgGlnGlnValProAspTyrSerLeuMet 99

 214 TCGGAGCCATTCGTGTGCTGTTTACTGTGTAGCGCTTGATGACTTCCTTG 263
     :::                             :: ||
 100 AlaIlePpMetValThrLeuCys....................Cy 107

 264 CTGTACTCCGCTTGGAGCCATTTGAGTGCGTGTTTACCGTGCCAGCACTCGA 313
     |||||  :::|||  ::::|||   :::||||
 107 sCysLeuProPheGlyLeuVal...AlaValPheMetAlaSerArgAlaL 123

 314 AGACTTCCTTGCTGTAC................TCGCTGGCAGCATAGTCTGTT 351
     ::|||     :||||                 ||||||::|||
 123 ysAspLysGlnMetTyrGlyAspAspAspGlyAlaArgSerThrSerGln 139

 352 TGCTGCGTCGGTCTGG.........GGACACCTGTGTACTACAGTACTTCG 392
     |||::|||  ||||            |||||||:||||
 140 CysAlaLeuGlyTrpAlaValAlaGlyIleValCysGlyIleValLeuSe 156

 393 TTTCCCTTTGTAGTGGTACTACTTC 417
     ::: |||||||||||
 156 rIleLeuLeuGlyIleTyrTyrPhe 164
```

seq_name: sp_organelle:O99381

seq_documentation_block:
```
ID O99381         PRELIMINARY;    PRT;   353 AA.
AC O99381;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COXI.
OS Pseudochorda nagaii.
OG Mitochondrion.
OC Eukaryota; stramenopiles; Phaeophyceae; Laminariales;
OC Pseudochordaceae; Pseudochorda.
OX NCBI_TaxID=74379;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEMURO, HOKKAIDO, JAPAN;
RA Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA Ohama T.;
RT "Phylogenetic analyses of heterokont algae with emphasis on the AUA
RT codon reassignment in mitochondria.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC     CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC     3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC     CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC     CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC     AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC     AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC     FERRICYTOCHROME C.
CC -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AF037992; AAC94978.1; -.
DR InterPro; IPR000883; -.
DR InterPro; IPR002106; -.
```

```
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER      1      1
FT   NON_TER    353    353
SQ   SEQUENCE   353 AA;  38157 MW;  062CBE2586F678DE CRC64;

alignment_scores:
            Quality:   78.00              Length:   145
              Ratio:    1.040               Gaps:     8
Percent Similarity:   51.724   Percent Identity:   26.897

alignment_block:
US-09-540-234-1 x O99381

Align seg 1/1   to: O99381   from: 1  to: 353

 92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGTCGTCTC 141
    ||||||||||::: ||| ::  |||  |||
109 AlaGlyThrGly...TrpThrValTyrProLeuSerGlyIleGlnAl 124

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAAC 191
    ::||| |||         :::           :::
124 aHisSer......GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGT 241
    |||||||||||||:::: |||||||||||||||:: ::: |||:: |||:
139 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrIl 155

242 GTAGCGCCTTG................ATGACTTCCTTGCTGCTACTCCG 273
    : :::::                    ||| ||| |||
155 ePheAsnMetArgSerProGlyMetThrMetGluArgLeuProLeuPheV 172

274 CTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    :||||| |||  :::::::| |||        :::|
172 alTrpSerValLeuIleThrAlaPheLeuLeuLeuSer......Leu 186

324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG........... 356
    |||||| :::|||||| :::||||||| :::
187 ProValLeuAlaGlyLeuThrLeuThrLeuThrLeuThrAspArgAsnPheAs 203

357 ..............CCTTTGTAGTGGTACTACTTCTT... 387
    ||| |||  ||:::::||||||::||
203 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 236

388 CTTCGTT...........CCCTTTGTAGTGGTACTACTTCTT... 419
    |||||| |||    |||::||||||::||

420 ...............AGCCATGTACTCGTCGT 437
    ||||||:::|||::|
237 GlyPheGlyIleIleValSerHisIleLeuAlaThr 247
```

seq_name: sp_organelle:O99382

seq_documentation_block:

CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

```
DR   EMBL; AF037993; AAC94979.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER      1      1
FT   NON_TER    353    353
SQ   SEQUENCE   353 AA;  38081 MW;  7D0B587DEF67DE0C CRC64;

alignment_scores:
            Quality:   78.00              Length:   143
              Ratio:    1.068               Gaps:     8
Percent Similarity:   51.049   Percent Identity:   51.049

alignment_block:
US-09-540-234-1 x O99382

Align seg 1/1   to: O99382   from: 1  to: 353

 92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGTCGTCTC 141
    ||||||||||::: ||| ::: ||| |||
109 AlaGlyThrGly...TrpThrValTyrProLeuSerGlyIleGlnAl 124

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAAC 191
    ::||| |||         :::           :::
124 aHisSer......GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGT 241
    |||||||||||||:::: |||||||||||||||:: ::: |||:: |||:
139 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrIl 155

242 GTAGCGCCTTG................ATGACTTCCTTGCTGCTACTCCG 273
    : :::::                    ||| ||| |||
155 ePheAsnMetArgAlaProGlyMetThrMetAspArgLeuProLeuPheV 172

274 CTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    :||||| |||  :::::::| |||        :::|
172 alTrpSerAlaLeuIleThrAlaPheLeuLeuLeuSer......Leu 186

324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG........... 356
    |||||| :::|||||| :::||||||| :::
187 ProValLeuAlaGlyAlaValThrMetLeuThrAspArgAsnPheAs 203

357 ..............CCTGGTCTGGGACACCTGTGTACTACAGTA 387
    ||||||| ||::::|
203 nThrThrPheAspProAlaGlyLyGlyAspProValLeuTyrGlnH 220

388 CTTCGTTT...........CCCTTTGTAGTGGTACTACTTCTT... 419
```

```
113 rHis.....AlaGlyAlaSerVal.AspLeuAlaIlePheSerLeuHis 127
         |||                 :::|||::|||
192 AACGGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGCTGTTACTGT 241
    |||::::     :::::::||:::||||:  :||    ::: :||
128 LeuAlaGlyAlaSerSerIleLeuGlyAlaValAlaAsnPheIleThrSerVa 144
242 GTAGCGCTTGATGACTTCC...........TTGCTGTACTCCG 273
         |||  ||||||                |||
144 IleAsnLeuArgThrSerGlyLeuThrMetGluArgLeuProLeuLeuV 161
274 CTTGGAGCCATTGAGTGTGTTACCGTGCAGCACTGCAAGACTTCCTT 323
    :::|||     |||:::||    ||||||      :::
161 alTrpAlaValLysIleThrValIleLeuLeuLeuLeuSer.....Leu 175
324 GCTGTACTCGCTGGCAGCATAGTCGTTGCTG...........356
    ||||||||||||     :::|||:::       ||||||
176 ProValLeuAlaGlyValAlaIleThrMetLeuLeuThrAspArgAsnPheAs 192
357 ...........CCTGGTCTGGGACACCTGTACTACAGTA 387
                     ||| |||  |||||| :::::
192 nThrSerPhePheAspProAlaGlyGlyGlyAspProValLeuPheGlnH 209
388 CTTCGTTT...........CCCTTGTAGTGTACTACTTCTT... 419
         |||        |||   |||   ||||||:::::::
209 IleSerLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 225
420 ...........AGCCATGTACTCGTGTCT 437
                   ||||||:::::::::|||
226 GlyPheGlyIleIleSerHisIleValThrSer 236
```

seq_name: sp_organelle:Q9T9P4

seq_documentation_block:
ID   Q9T9P4   PRELIMINARY;   PRT;   513 AA.
AC   Q9T9P4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Terebratulina retusa.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Brachiopoda; Rhynchonelliformea; Rhynchonellata;
OC   Terebratulidina; Cancellothyroidea; Cancellothyrididae; Terebratulina.
OX   NCBI_TaxID=7580;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stechmann A., Schlegel M.;
RT   "Analysis of the complete mitochondrial DNA sequence of the brachiopod
RT   Terebratulina retusa places Brachiopoda within the protostomes.";
RL   Proc. R. Soc. Lond., B, Biol. Sci. 260:2043-2053(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AJ245743; CAB59842.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR001505; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   ProDom; PD000131; -; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.

```
220 IleLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 236
         |||                 |||       :::|||::::|||
420 ...........AGCCATGTACTC 431
                   |||||||:::|||
237 GlyPheGlyIleValSerHisIleLeu 245
```

seq_name: sp_organelle:Q9MJ87

seq_documentation_block:
ID   Q9MJ87   PRELIMINARY;   PRT;   406 AA.
AC   Q9MJ87;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COI.
OS   Terebratulina pacifica.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Brachiopoda; Rhynchonelliformea; Rhynchonellata;
OC   Terebratulidina; Cancellothyroidea; Cancellothyrididae; Terebratulina.
OX   NCBI_TaxID=93899;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TPC;
RX   MEDLINE=20318514; PubMed=10860643;
RA   Saito M., Kojima S., Endo K.;
RT   "Mitochondrial COI sequences of brachiopods: genetic code shared with
RT   protostomes and limits of utility for phylogenetic reconstruction.";
RL   Mol. Phylogenet. Evol. 15:331-344(2000).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB026515; BAB08018.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER   1   1
FT   NON_TER   406   406
SQ   SEQUENCE   406 AA; 44252 MW; 6EB3A91331272286 CRC64;

alignment_scores:
        Quality: 78.00   Length: 145
        Ratio: 1.040   Gaps: 8
Percent Similarity: 51.724   Percent Identity: 27.586

alignment_block:
US-09-540-234-1 x Q9MJ87

Align seg 1/1   to: Q9MJ87   from: 1   to: 406

```
 92 GCTTGGCACGGCCATAATGGAAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTC 141
    |::::  |||  :::|||  :::||| :::||   |||||||||| |||:::||
 98 AlaGlyThrGly...TrpThrValTyrProProLeuAlaAlaIleThrSe 113
142 AAACATGCCCCCTCGAGACGTCCGTGAAGCCGTAGTTTGTACAAGACTAAC 191
    |:::     |::||||| ::: :::||||||     :::::::::    :::
```

KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
SQ   SEQUENCE   513 AA;   56600 MW;   CBIF2D0F1A4CBC08 CRC64;

alignment_scores:
         Quality: 78.00   Length: 145
           Ratio: 1.040   Gaps: 8
Percent Similarity: 51.724   Percent Identity: 27.586

alignment_block:
US-09-540-234-1 x Q9T9P4   ..

Align seg 1/1 to: Q9T9P4 from: 1 to: 513

   92 GCTGGCACGGCATAATGGAAAACAGTCCTCGCCCTTTGAGGGTCGTCTC 141
      |||||||||::: ||| :: ||| ||| :::::||
  119 AlaGlyThrGly...TrpThrValTyrProProLeuAlaSerIleThrSe 134

  142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAAC 191
      |:::   :::||||| ::: ::: ||||||:: ::
  134 rHis.....AlaGlyAlaSerVal.AspLeuAlaIlePheSerLeuHis 148

  192 AACGGCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGT 241
      |||:::::: :::::::::::||||||::::: ::: |||||| :||
  149 LeuAlaGlyAlaSerSerIleLeuGlyAlaValAsnPheIleThrSerVa 165

  242 GTAGCGCCTTGATGACTTCC..............TTGCTGTACTCCG 273
      ||:::||| |||||| |||
  165 lIleAsnLeuArgThrSerGlyLeuThrMetGluArgLeuProLeuLeuV 182

  274 CTTGGAGCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
      ::|||:   :||| |||
  182 aIThrpAlaValLysIleThrVallIleLeuLeuLeuLeuSer......Leu 196

  324 GCTGTACTCCGCTGGCAGCAATAGTCGTCTGTTGCTG.............. 356
      |||||||||||:::||||::: |||||
  197 ProValLeuAlaGlyValAlaIleThrMetLeuLeuThrAspArgAsnPheAs 213

  357 ..........CCTGGCTGGGACACCTGTGTACTACAGTA 387
                      ||| ||| ||||||| :::::
  213 nThrSerPheAspProAlaGlyGlyGlyAspProValLeuPheGlnH 230

  388 CTTCGTT..............CCCTTGTAGTGTACTACTTCTT.... 419
      || ::::||| || |||||||||:::||:::
  230 lsLeuPheTrpPheGlyHisProGluValTyrIleLeuIleLeuPro 246

  420 ...............AGCCATGTACTCCGTGCT 437
                         ||||||||:::|||
  247 GlyPheGlyIleIleSerHisIleValThrSer 257

seq_name: sp_organelle:O21002

seq_documentation_block:
ID   O21002   PRELIMINARY;   PRT;   515 AA.
AC   O21002; O47422;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Branchiostoma lanceolatum (Common lancelet) (Amphioxus), and
OS   Branchiostoma floridae (Florida lancelet) (Amphioxus).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC   Branchiostoma.
OX   NCBI_TaxID=7740, 7739;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.lanceolatum;
RX   MEDLINE=98292550; PubMed=9628930;
RA   Spruyt N., Delarbre C., Gachelin G., Laudet V.;
RT   "Complete sequence of the amphioxus (Branchiostoma lanceolatum)

RT   mitochondrial genome: relations to vertebrates.";
RL   Nucleic Acids Res. 26:3279-3285(1998).
RN   [2]
RP   SEQUENCE OF 1-8 FROM N.A.
RC   SPECIES=B.lanceolatum;
RX   MEDLINE=97398704; PubMed=9254918;
RA   Delarbre C., Barriel V., Tiller S., Janvier P., Gachelin G.;
RT   "The main features of the craniate mitochondrial DNA between the ND1
RT   and the COI genes were established in the common ancestor with the
RT   lancelet.";
RL   Mol. Biol. Evol. 14:807-813(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.floridae;
RA   Naylor G.J.P., Brown W.M.;
RT   "Amphioxus mitochondrial DNA, chordate phylogeny, and the limits of
RT   inference based on comparisons of sequences.";
RL   Syst. Biol. 47:61-76(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.floridae;
RX   MEDLINE=97394466; PubMed=9252184;
RA   Naylor G.J., Brown W.M.;
RT   "Structural biology and phylogenetic estimation.";
RL   Nature 388:527-528(1997).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   SPECIES=B.floridae;
RX   MEDLINE=99261652; PubMed=10331267;
RA   Boore J.L., Daehler L.L., Brown W.M.;
RT   "Complete sequence, gene arrangement, and genetic code of
RT   mitochondrial DNA of the cephalochordate Branchiostoma floridae
RT   (Amphioxus).";
RL   Mol. Biol. Evol. 16:410-418(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; Y16474; CAA76249.1; -.
DR   EMBL; Y09524; CAA70710.1; -.
DR   EMBL; AF098298; AAB87991.2; -.
DR   HSSP; P98002; 1AR1.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1_UNKNOWN_1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
SQ   SEQUENCE   515 AA;   57459 MW;   F915C249B0C83767 CRC64;

alignment_scores:
         Quality: 78.00   Length: 115
           Ratio: 1.300   Gaps: 5
Percent Similarity: 52.174   Percent Identity: 25.217

alignment_block:
US-09-540-234-1 x O21002   ..

Align seg 1/1 to: O21002 from: 1 to: 515

  174 AGTTGTACAAGACTAACAACGGCCAATGGCGTCGATGTTATCGGAGCCAT 223
      :: ::: ||| ::: |||||:::||||||
  146 AlaIlePheSerLeuHisLeuAlaGlyValSerSerIleLeuGlyAlaIl 162

```
224 TCGTGTGCTGTTACTGTGTAGCGCTTGATGACTTCCTTGCTGCTACTCC..  272
        |||| :::  |||:  |||::   |::::  :||||::  :|:::
162 eAsnPheIleThrThrIleHisAsnMetArgAlaSerIleGluTrpAsnA  179

273 .........GCTTGGAGCCATTGAGTCGTGTTACCGTGCAGCA  308
              ::|||||      ::|||||||||:||
179 rgValProLeuPheValTrpSerIleTrpValThrAlaTyrLeuLeuLeu  195

309 CTCGAAGACTTCCTTGCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG..  356
    |||:::  |||  |||||||||:::||:|  ||||||
196 Leuser....LeuProValLeuAlaGlyValAlaIleThrMetLeuLeuTh  210

357 .......................CCTGGTCTCGGGGACAC  372
                           |||  ||| |||
210 rAspArgAsnIleAsnThrThrPhePheAspProSerGlyGlyGlyAspP  227

373 CTGTGTACTACAGTACTTCGTTT...............CCCTTTGTAGTG  407
    ||:: |||::: |||          |||   |||
227 roIleLeuTyrGluHisLeuPheTrpPhePheGlyHisProGluValTyr  243

408 GTACTACTTCTT...............AGCCATGTACTCGTG  434
    ::||::::||                |||||||:::::||||
244 IleLeuIleLeuProGlyPheGlyIleIleSerHisIleIleIle  258

seq_name: sp_plant:Q96397
```

```
alignment_scores:
    Quality:     78.00           Length:        192
      Ratio:      1.099            Gaps:          8
Percent Similarity:   36.979     Percent Identity:  21.875

alignment_block:
US-09-540-234-1 x Q96397

Align seg 1/1  to: Q96397  from: 1  to: 640

 11 GCCCTTATGTATGCAGGAGCGCGCCTAGTGGTCATCGAAGGAAAGGACAA  60
       |||||||:||   |||||::|| |||||||    |||||
427 AlaAlaValProAlaAlaAlaCysSerGlyPheArgGlyGlyArgGlyGl  443

 61 AGACACCTCCAAGCCGGACACGGCTACTGTAGCTGGCACGGCATAATGGA 110
    :::|||::|||||||||::          |||||||||||||:::
443 yGlyValAlaArg........ValAlaAlaGlyThrSerArgAlaA  456

111 AAACAGTCCCTGCCCTTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACG 160
```

```
456 laGlyValProArgArgLeu...........................  462
    |||||
161 TCGGTGAAGCCGTAGTTGTACAAGACTAACAACGCCAATGGCGTCGA..  208
    ||||||:::|||||||||
463 ...........................GlnArgArgTrpArgArgAr  469

208 ...........................GlnArgArgTrpArgArgAr  208
469 gGlyArgGlyTrpArgArgValArgArgArgArgValAlaGlyArgA  486

209 ....TGTTATCGGAGCCATTCGTGCTGTTACTGTGTAGCGCTTGAT  253
        ||| :::  |||  :::  |||:::
486 laValCysThrAlaGlyArgCysCysTrpMetThrCysLeuProMetTrp 502

254 GACTTCCTTGCTGCTACTCCGCTTGGAGCCATTGAG............  289
    ::|||||
503 GlySerGlyGlyThrTrpProTrpArgProLeuMetThrProSerArgTh 519

290 ....TGCTGTTACCGTGCAGCACTCGAAGACTTCCTTGCTGTACT.....  331
    ::|||||||||
519 rCysAlaCysLeuPro.....................ThrProCysSerSerArgT 531

332 ...................................CGCTGGCAGCAT  343
                                       ||||||:::::
531 rpLeuArgArgTrpArgCysGlyTrpAlaProGlyGlyArgTrpArgCys 547

344 AGTTCTGTTTGCTGCCTGG..............TCTGGGGACACCTGTGTACT 381
    ||||||  :::::|||        ||||||  |||  :::||
548 SerLeuCysSerCysTrpArgTrpGlyCysSerGlyArgGlyThrProLeuLe 564

382 ACAGTACTTCGTT...................TCCCTTTGTAGTGGTACTA 413
    |  ::|||                    |||  :::
564 uProThrTrpValTrpArgArgArgCysCysArgCysCysArgGlySerA 581

414 CTTCTTAGCCATGTACTACTCGTGTCTGT 439
    :::  |||:::  |||||||
581 rgAlaProArgCysAsnTrpValCys 589
```

seq_name: sp_invertebrate:O45936

```
DR   EMBL; AL021481; CAA16336.1; -.
DR   InterPro; IPR002422; -.
DR   Pfam; PF01490; Aa_trans; 1.
SQ   SEQUENCE   607 AA;  68571 MW;  8C643382F7AED090 CRC64;

alignment_scores:
        Quality:     77.50           Length:    136
          Ratio:      1.062            Gaps:      8
Percent Similarity:  53.676   Percent Identity:  24.265

alignment_block:
US-09-540-234-1 x 045936

Align seg 1/1 to: 045936  from: 1 to: 607

 58 CAAAGACACCTCCAAGCCGGACACGGCTACTGTAGCTGGCACGGCATAAT 107
    :::::::::::::::::::|||  ||| :||||||:::  ||||:||||||
266 GluAsnArgMetGlnSerProHisAlaPheIleSerTrpAsnGlyValIle 282

108 GGAAAACAGTCCCTGC.........................CCTT 127
    :  |||||  |||  ||                         ||||
282 u...AsnSerCysLeuValValLeuAlaIleLeuPheSerValThrGlyP 298

128 TTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTT 177
    ||  || ||| |||:   ::||||||         ::|||||::::::
298 heTyrGlyTyrLeuSer...LeuGlyAsnAspValLysAspThrAlaThr 313

178 TGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCATTCGT 227
     :: ||||||||:::|||  ::|||: |:|||||:||:||:::||||||
314 LeuAsnLeuProMetThrProPheTyrGlnThrIleLysLeuMetPheVa 330

228 GTGCTGTTTACTGTGTAGCG......CTTGATGACTTCCTTGCTGCTACTC 271
    |:::::||||: :::|||  :::           ||| |||||||||::: ::
330 lAlaCysIleMet.IleSerTyrProLeuGlnPheTyrValProMetGlu 346

272 CGCTTGGAGCCATTTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCC 321
    ||:::||| :::                          |||||:::::||
347 ArgIleGluLysTrp...............IleThrArgLysIlePr 357

322 TTGCTGTACTCGCTGGCAGCATAGTCGTGTTGCTGCCTGGTCTGGGGAC. 370
     :: |||||| ::  |||||| ||||:|||:|  ::|||:|:|||||||
357 oValAspLysGlnThrLeuTyrIleTyrIleAlaAargTyrSerGlyValI 374

371 ....ACCTGTGTA................CTACAGTACTTCGTT 394
    ||||||:::||                  ||||||:::::||
374 leLeuThrCysAlaIleAlaAlaGluLeuIleProHisLeuAlaLeuPheIle 390

395 TCCCTT 400
    ||||||
391 SerLeu 392

seq_name: sp_virus:090368

seq_documentation_block:
ID   090368     PRELIMINARY;     PRT;    2513 AA.
AC   090368;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   NONSTRUCTURAL POLYPROTEIN.
OS   O'nyong-nyong virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC   Alphavirus.
OX   NCBI_TaxID=11027;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SG650;
RA   Lanciotti R.S., Ludwig M.L., Rwanuma E.B., Lutwama J.J., Kram T.M.,
RA   Karabatsos N., Cropp B.C., Miller B.R.;
RT   "Emergence of Epidemic O'nyong-nyong Fever in Uganda After a 35 Year
RT   Absence:Genetic Characterization of the Virus.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
```

```
DR   EMBL; AF079456; AAC97204.1; -.
DR   InterPro; IPR000606; -.
DR   InterPro; IPR001788; -.
DR   InterPro; IPR001789; -.
DR   InterPro; IPR002589; -.
DR   InterPro; IPR002620; -.
DR   Pfam; PF00978; RNA_dep_RNApol2; 2.
DR   Pfam; PF01443; Viral_helicase1; 1.
DR   Pfam; PF01661; DUF27; 1.
DR   Pfam; PF01707; Peptidase_C9; 1.
DR   PROSITE; PS00093; N4_MTASE; UNKNOWN_1.
KW   Polyprotein.
FT   CHAIN          1    535       NON-STRUCTURAL PROTEIN 1.
FT   CHAIN        536   1333       NON-STRUCTURAL PROTEIN 2.
FT   CHAIN       1334   1896       NON-STRUCTURAL PROTEIN 3.
FT   CHAIN       1897   2513       NON-STRUCTURAL PROTEIN 4.
SQ   SEQUENCE   2513 AA;  280219 MW;  1B546AD981E5F2A7 CRC64;

alignment_scores:
        Quality:     77.50           Length:    163
          Ratio:      0.912            Gaps:     11
Percent Similarity:  52.147   Percent Identity:  29.448

alignment_block:
US-09-540-234-1/rev x 090368

Align seg 1/1 to: 090368  from: 1 to: 2513

 417 GAAGTAGTACCACTACAAAGGAAAACGAAGTACTGT........... 382
     |||:|:||||::       ::: :|:        |||
1686 GluAlaValProSerAspLeuGluAlaCysAspAlaThrMetAs 1702

 381 ......AGTACCAGGTGTCCCAGACCAGGCAGCAAACAGACTATGCT 339
     |||     |||  ::: ||| ||||||||||:::::||
1702 pTrpProSerIleAspIleValPro.ThrArgGlnArgSerAspSerPhe 1718

 338 GCCAGCGAGTAC...AGCAAGGAAGTCTTCGAGTGCTGCACCGTAAAC.. 294
     ||||||||  :::          :: |||||:::::||||||  :::
1719 AspSerGluTyrSerSerArgSerAsnIleGlnLeuValThrAlaAspVa 1735

 293 ...........AGCACTCAAATGGCTCCAAGCGGAGTACAGCAAG 260
     :::::::::::||||||| |||:::||||||||
1735 lHisAlaProMetTyrAlaAsnSerLeuAlaSerGly........G 1749

 259 GAAGTCATCAAGCGCTACACAGTAAACACACGAATGGC........... 219
     |||||    ::::|||  :::||| ||||||||||||
1749 lySerValLeuSerLeuSerGluGlnAlaGlnAsnGlyIleMetIle 1765

 218 ...........TCCGATAACATCGACGCCATTGGCGTTGTTAGTCTTGTACA 178
     |||||||||||||:::  ||||:::  ::: ||||||||||
1766 LeuProAspSerGluAspThrAspSerIle........SerArgValSe 1779

 177 AACTACGGCTTCACCGACGTCTCGAGGG.................G 149
     :|||||||||||:::                       |||
1779 rThrProIleAlaProProArgArgLeuGlyArgThrIleAsnValT 1796

 148 CATGTTTGAGACGACCCTCAAAAGG.........GCAGGGACTGTTTC 109
     ::||| :::||  ::|||        :::::          |||:::
1796 hrCysAspGluArgGluGlyLysIleLeuProMetAlaSerAspArgLeu 1812

 108 CATTATGCCGTGCCAGCTACAGTAGCC.......GTGTCCGGCTTGGAGGT 65
     |||||||||||||||||:::::         |||||||||||||
1813 PheThrAlaLysProTyrThrValAlaLeuGlyValSerThrAlaAspIl 1829

  64 GTCTTTGTCCTTTCCTTCAGATGACCCACTAGGCGCT 28
     :::  :::||||||||  :: |||||||||||||:::
1829 eThr...AlaTyrProGlnAlaProLeuGlySer 1840

seq_name: sp_organelle:099385

seq_documentation_block:
ID   099385     PRELIMINARY;     PRT;     353 AA.
```

```
AC   O99385;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Tribonema marinum.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Xanthophyceae; Tribonematales;
OC   Tribonemataceae; Tribonema.
OX   NCBI_TaxID=74375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIES 548;
RA   Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA   Ohama T.;
RT   "Phylogenetic analyses of heterokont algae with emphasis on the AUA
RT   codon reassignment in mitochondria.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AF037996; AAC94982.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER        1      1
FT   NON_TER      353    353
SQ   SEQUENCE   353 AA;  38121 MW;  7DDCDED749BA4681 CRC64;
```

alignment_scores:
                 Quality: 77.00   Length: 143
                 Ratio: 1.055   Gaps: 8
Percent Similarity: 51.049   Percent Identity: 27.273

alignment_block:
US-09-540-234-1 x O99385

Align seg 1/1 to: O99385 from: 1 to: 353

```
 92  GCTGGCACGGCATAATGGAAAACACTCCCTGCCCTTTTGAGGGTCGTCTC  141
     ||||||||||:   |||  ::|  |||  |||   :: |||::
109  AlaGlyThrGly...TrpThrValTyrProLeuSerGlyValGlnIleGlnIleAl  124

142  AAACATGCCCCTCCGAGACCGTCGGTGAAGCCGTAGTTGTACAAGACTAAC  191
     ::||:      |||||| ||||: :::::::|||  :::
124  aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis  138

192  AACGCCAATGGCGTCGATGTTATCGGAGGCCATTCGTGTCGTGTTACTGT  241
     ::|:|||||||:||||||||||||||||:::|| |||::    |||::
139  LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrIl  155

242  GTAGGCGCTTG..........ATGACTTCCTTGCTGTACTCCG  273
     :  :||||||                 |||        |||
155  ePheAsnMetArgAlaProGlyMetThrMetAspArgLeuProLeuPheV  172

274  CTTGGAGCCATTTGAGTGCTGTGTTTACCGTGCAGCACTCGAAGACTTCCTT  323
```

```
                                        ::|||||||   |||     ::::::::::   |||::           |||
172  alTrpSerValLeuIleThrAlaPheLeuLeuLeuLeuSer......Leu  186

324  GCTGTACTCGCTGGCAGCATAGTCTGTTGCTG.................  356
     ||||||||||||||||:::|||::  ||||||
187  ProValLeuAlaGlyGlyIleThrMetLeuThrAspArgAsnPheAs  203

357  ...........CCTGGTCTGGGACACCTGTTACTACAGTA  387
                ||  ||| ||||||||||||||||  |||::
203  nThrThrPheAspProAlaGlyGlyAspProValLeuTyrGlnH  220

388  CTTCGTTT............CCCTTTGTAGTGGTACTACTTCCTT...  419
     |||              ||| |||    :::||||:::|||::|||
220  isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro  236

420  ..............AGCCATGTACTC  431
                   ||||||::::|||
237  GlyPheGlyIleValSerHisIleLeu  245
```

seq_name: sp_organelle:O47467

seq_documentation_block:
```
ID   O47467         PRELIMINARY;      PRT;     353 AA.
AC   O47467;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Heterococcus caespitosus.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Xanthophyceae; Heterococcus.
OX   NCBI_TaxID=55585;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 835/2A;
RX   MEDLINE=97383241; PubMed=9236270;
RA   Ehara M., Hayashi-Ishimaru Y., Inagaki Y., Ohama T.;
RT   "Use of a deviant mitochondrial genetic code in yellow-green algae as
RT   a landmark for segregating members within the phylum.";
RL   J. Mol. Evol. 45:119-124(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 835/2A;
RX   MEDLINE=97451042; PubMed=9302324;
RA   Inagaki Y., Hayashi-Ishimaru Y., Ehara M., Igarashi I., Ohama T.;
RT   "Algae or protozoa: phylogenetic position of euglenophytes and
RT   dinoflagellates as inferred from mitochondrial sequences.";
RL   J. Mol. Evol. 45:295-300(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 835/2A;
RX   MEDLINE=98004465; PubMed=9342410;
RA   Hayashi-Ishimaru Y., Ehara M., Inagaki Y., Ohama T.;
RT   "A deviant mitochondrial genetic code in prymnesiophytes (yellow-
RT   algae): UGA codon for tryptophan.";
RL   Curr. Genet. 32:296-299(1997).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
EMBL; AB000206; BAA24970.2; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR000894; -.
```

```
DR   Pfam; PF00115; COX1. 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   ProDom; PD000290; -; 1.
DR   PROSITE; PS00077; COX1. 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER     1     1
FT   NON_TER   353   353
SQ   SEQUENCE   353 AA;   38214 MW;   3FA28DC6FFAF2FEA CRC64;

alignment_scores:
         Quality:     77.00              Length:      145
         Ratio:      1.041                Gaps:          9
Percent Similarity:  51.034   Percent Identity:  28.276

alignment_block:
US-09-540-234-1 x 047467    .

Align seg 1/1  to: 047467   from: 1 to: 353

92  GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    |||||||||| |||  ::: || ||| |||||:::  |||||||::
109 AlaGlyThrGly..TrpThrValTyrProLeuSerSerValValAl 124

142 AAACATGCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    :::          |||||| ::: ||| |||||||:::
124 aHisSer......GlyProSerVal AspLeuAlaIlePheSerLeuHis 138

192 AACGGCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTCTGTTTACTGT 241
    ::::::::::||||||||:::  ||| ::: |||
139 LeuSerGlyAlaSerSerIleLeuGlyAlaIleLeuAsnPheIleThrThrIl 155

242 GTAGCGCTTG................ATGACTTCCTTGCTGTACTCCG 273
       :::::                  |||  |||
155 eGlnAsnMetArgAlaProGlyMetSerMetHisArgLeuProLeuPheV 172

274 CTTGGAGCCATTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTT 323
    :::          |||||||||::::: :::|||
172 alTrp..........AlaIleTyrIleThrAlaPheLeuLeuLeuLeu 184

324 GCT....GTACTCGCTGGCAGCAGCATAGTCTGTTTGCTG.......... 356
    |||       |||||||||:::||| |||
185 AlaLeuProValLeuAlaGlyAlaIleThrMetLeuLeuThrArgAspArgAs 201

357 ...............CCTGGTCGGGGACACCTGTACT 381
                   ||| |||  ||||||:::: |
201 nPheAsnThrSerPhePheAspProAlaGlyGlyGlyAspProIleLeuT 218

382 ACAGTACTTCGTTT............CCCTTGTAGTGGTACTACTT 416
    |||::|            ||| |||:::||:::
218 yrGlnHisLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIle 234

417 CTT...............AGCCATGTACTC 431
    |||
235 LeuProGlyPheGlyIleIleSerHisIleLeu 245


seq_name: sp_plant:Q9LJ36

seq_documentation_block:
ID   Q9LJ36       PRELIMINARY;    PRT;    474 AA.
AC   Q9LJ36;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL PROTEIN.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC   Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN-CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0025D05.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001072; BAA89546.1; -.
DR   InterPro; IPR001623; -.
DR   Pfam; PF00226; DnaJ; 1.
DR   PROSITE; PS50076; DNAJ_2; 1.
DR   SMART; SM00271; DnaJ; 1.
SQ   SEQUENCE   474 AA;   51275 MW;   97E9CF9F70291705 CRC64;

alignment_scores:
         Quality:     77.00              Length:      128
         Ratio:      1.013                Gaps:          5
Percent Similarity:  59.375   Percent Identity:  25.000

alignment_block:
US-09-540-234-1 x Q9LJ36    ..

Align seg 1/1  to: Q9LJ36   from: 1 to: 474

2   GGTCCACGGCCCTTATGTATGCAGGAGCCGCCTAGTGGTCATCTGAAGG 51
    |||          |||:::::|||  ::: :::|||::::
142 GlySerTyrIleLeuPheAlaLeuGlyLysGlyHisGlyHisAl 158

52  AAAGGACAAAGACACCTCCAAGCCGGACACGGCTACTGTAGCTGGCACGG 101
    :::         |||||||||          ::|||||||||
158 aHisAsnHisProMetGluLys...........MetAlaValAlaGlyLeuv 172

102 CATAATGGAAAACATGCCTGCCCTTTGAGGGTCGTCAAACATGCCC 151
    ||||||||||  ||||||
172 al........LeuValProAlaLeuSerProCysAlaThrThrLeuPro 185

152 CTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATG 201
    :::          ||||:::::|||:::
186 Val...........PheLeu.AlaValGlyAsnSerSers 195

202 GGCTGCGATGTTATCGGAGCCATTCGTCGTGTTGTTACTGTGTAGCGCTTG 251
    |||          ::: :||||||:::|||
195 erMetMetIleLeuAlaIleIleValLeuLeuPheSerThrIleThrVal 211

252 ATGACTTCCTTGCTGTACTCCCGTTGGAGCCATTGAGTGCTGTTTACCG 301
    ||||||:::|||||||
212 MetThrSerLeuValAlaLeuSerPheTyrGlyAlaSerGlnIleLeuLysPh 228

302 TGCAGCACTCGAAGACTTCCTTGCTGTACTCGCTGGCAGCCATAGTCTGTT 351
    :::          ||||:::::|||:::::::
228 eHisTrpValGluArgTyrAspLysIleLeuValGlyServalLeuCysL 245

352 TGCTGCCTGGTCTGGGGACACCTGTACTAC 383
    |||          ||| |||  |||:::
245 euVal...........GlyValLeuThrTyrValPheHis 254


seq_name: sp_organelle:P92800

seq_documentation_block:
ID   P92800       PRELIMINARY;    PRT;    528 AA.
AC   P92800;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1.
OS   Pylaiella littoralis.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales;
OC   Acinetosporaceae; Pylaiella.
OX   NCBI_TaxID=2885;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Fontaine J., Goux D., Kloareg B., Loiseaux-de Goer S.;
```

```
RL   J. Mol. Evol. 0:0-0(0).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; 272500; CAA96581.1; -.
DR   HSSP; P98002; 1AR1.
DR   Mendel; 21583; Pylll;cox1;21583.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
SQ   SEQUENCE 528 AA; 58015 MW; 11E3DDBD9A055BAE CRC64;

alignment_scores:
        Quality:     77.00           Length:     143
        Ratio:        1.055          Gaps:         8
Percent Similarity:  51.049     Percent Identity: 27.273

alignment_block:
US-09-540-234-1 x P92800

Align seg 1/1  to: P92800  from: 1  to: 528

 92 GCTGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    |||||||||::: ||| ||:::||  ||  ||||||:::
149 AlaGlyThrGly...TrpThrValTyrProLeuSerGlyIleGlnAl 164

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    ::::            |||||  :::::::::::::::::  :::
164 aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 178

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGTTTACTGT 241
    :::::::::::::::::|||||||||  :: |||:::       |||:
179 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrThrIl 195

242 GTAGGCGCTTG..........ATGACTTCCTTGCTGTACTCCG 273
    :::::::            |||  ||            :
195 ePheAsnMetArgAlaProGlyMetThrMetAspArgLeuProLeuPheV 212

274 CTTGGAGCCATTGAGTGCTGTGTTTACCGTGCAGCACTGTGACTACAGTA 323
    ||||||| |||  ||||||| |||:::  |||:::
212 alTrpSerValLeuIleThrAlaPheLeuLeuLeuSer.....Leu 226

324 GCTGTACTCGGCTGGCAGCATAGTCGTTGTTTGCTG............ 356
    |||||||||||:||:: :||||:::  |||||
227 ProValLeuAlaGlyLysIleThrMetLeuThrArgAspArgAsnPheAs 243

357 ..........CCTGGCTGGGACACCTGTGTACTACAGTA 387
    ||||  ||| ||||||| ||:::
243 nThrThrPhePheAspProAlaGlyGlyTyrGlyAspProValLeuTyrGlnH 260

388 CTTCGTTT..........CCCTTTGTAGTGGTACTACTTCTT... 419
    |||         |||:::|||:::
260 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 276

420 ..........AGCCATGTACTC 431
    ||||||||:::|||
277 GlyPheGlyIleValSerHisIleLeu 285
```

```
seq_name: sp_organelle:Q9TE66

seq_documentation_block:
ID   Q9TE66        PRELIMINARY;         PRT;       353 AA.
AC   Q9TE66;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Skeletonema costatum.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC   Thalassiosirophycidae; Thalassiosirales; Skeletonemataceae;
OC   Skeletonema.
OX   NCBI_TaxID=2843;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP 1077-1B;
RA   Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT   "Phylogenetic analysis of diatom coxI genes and implications of
RT   fluctuating GC content on mitochondrial genetic code evolution.";
RL   Curr. Genet. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB020227; BAA86612.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER    1       1
FT   NON_TER    353     353
SQ   SEQUENCE 353 AA; 38158 MW; F107054A59A65521 CRC64;

alignment_scores:
        Quality:     76.50           Length:     130
        Ratio:        1.142          Gaps:         5
Percent Similarity:  51.538     Percent Identity: 26.154

alignment_block:
US-09-540-234-1 x Q9TE66

Align seg 1/1  to: Q9TE66  from: 1  to: 353

135 TCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAA 184
    ||||||  |||||||||::: |||          :::  ::::::::
120 SerSerAlaThrAlaHisSerGlyGlySerValAspLeuAlaIlePheSe 136

185 GACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGCTGT 234
    ::|    :|||||:||:|:|||||||||:: |||
136 rLeuHisLeuSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleC 153

235 TTACTGTGTAGCGCTGATGACT............TCCTTGCTG 266
    :|||:: |||||||||:::||                 ::|||
153 ysThrIlePheAsnMetArgValLysSerLeuSerPheHisAsnLeuPro 169

267 TACTCCGCTTGGAGCCATTGAGTGCTGTGTTTACCGTGCAGCACTGCGAAGA 316
    :::||||||  |||              |||:::
```

```
 170 LeuPheValTrpSerValLeuIleThrAlaPheLeuLeuLeuSer.. 185

 317 CTTCCTTGCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG........ 356
     ||| ||||||||||||||||::::::||||||:    ||||||
 186 ....LeuProValLeuAlaGlyIleAlaIleThrMetLeuLeuThrAspArgA 201

 357 ................CCTGGTCTGGGGACACCTGTCTAC 380
                       ||| ||| |||||||
 201 snPheAsnThrPhePheAspProAlaGlyGlyGlyAspProValLeu 217

 381 TACAGTACTTCGTTT..................CCCTTTGTAGTGGTACTACT 415
     ::::::||   |||                    ||| ||   ||:||||
 218 PheGlnHisLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIl 234

 416 TCTT...............AGCCATGTACTCGTGTCT 437
     :|||               ||||||||::::||||||
 234 eLeuProGlyPheGlyIleValSerHisIleValValSer 247

seq_name: sp_organelle:Q9TE64

seq_documentation_block:
ID  Q9TE64        PRELIMINARY;       PRT;     353 AA.
AC  Q9TE64;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN  COXI.
OS  Thalassiosira nordenskioeldii.
OG  Mitochondrion.
OC  Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC  Thalassiosirophycidae; Thalassiosirales; Thalassiosiraceae;
OC  Thalassiosira;
OX  NCBI_TaxID=83372;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CCMP992;
RA  Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT  "Phylogenetic analysis of diatom coxI genes and implications of
RT  fluctuating GC content on mitochondrial genetic code evolution.";
RL  Curr. Genet. 0:0-0(1999).
CC  -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC      CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC      3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC      CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC      CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC      AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC      AND COPPER B (BY SIMILARITY).
CC  -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC      FERRICYTOCHROME C (BY SIMILARITY).
CC  -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC  -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR  EMBL; AB020229; BAA86614.1; -.
DR  InterPro; IPR000883; -.
DR  Pfam; PF00115; COX1; 1.
DR  PRINTS; PR01165; CYCOXIDASEI.
DR  PROSITE; PS00077; COX1; 1.
KW  Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW  Respiratory chain; Transmembrane.
FT  NON_TER       1      1
FT  NON_TER     353    353
SQ  SEQUENCE  353 AA;  38158 MW;  F107054A59A65521 CRC64;

alignment_scores:
         Quality:  76.50            Length:       130
           Ratio:  1.142              Gaps:         5
Percent Similarity:  51.538     Percent Identity:  26.154

alignment_block:
US-09-540-234-1 x Q9TE64
```

```
Align seg 1/1  to: Q9TE64  from: 1  to: 353

 135 TCGTCTCAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAA 184
     |||||| |||:| |||               ::::::::||| |
 120 SerSerAlaThrAlaHisSerGlyGlySerValAspLeuAlaIlePheSe 136

 185 GACTAACAACGCCAATGGCGTCGATGTTATCGGACCATTCGTGTGCTGT 234
     :|||||||||:|::|||||||||||||||||||||||||    :: ::
 136 rLeuHisLeuSerGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleC 153

 235 TTACTGTGTAGCCGCTTGATGACT.............TCCTTGCTG 266
     ||::|  |:::||::| ||: ::           ::::|||
 153 ysThrIlePheAsnMetArgValLysSerLeuSerPheHisAsnLeuPro 169

 267 TACTCCGCTTGGAGCCATTTGAGTGCTGTTACCGTGCAGCACTCGAAGA 316
     ::||||||   |||   :::::::||   ||| |||:|||
 170 LeuPheValTrpSerValLeuIleThrAlaPheLeuLeuLeuSer.. 185

 317 CTTCCTTGCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG........ 356
     ||| ||||||||||||||||::::::||||||:    ||||||
 186 ....LeuProValLeuAlaGlyIleAlaIleThrMetLeuLeuThrAspArgA 201

 357 ................CCTGGTCTGGGGACACCTGTCTAC 380
                       ||| ||| |||||||
 201 snPheAsnThrPhePheAspProAlaGlyGlyGlyAspProValLeu 217

 381 TACAGTACTTCGTTT..................CCCTTTGTAGTGGTACTACT 415
     ::::::||   |||                    ||| ||   ||:||||
 218 PheGlnHisLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIl 234

 416 TCTT...............AGCCATGTACTCGTGTCT 437
     :|||               ||||||||::::||||||
 234 eLeuProGlyPheGlyIleValSerHisIleValValSer 247

seq_name: sp_organelle:Q9G407

seq_documentation_block:
ID  Q9G407        PRELIMINARY;       PRT;     511 AA.
AC  Q9G407;
DT  01-MAR-2001 (TrEMBLrel. 16, Created)
DT  01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  CYTOCHROME C OXIDASE SUBUNIT 1.
GN  COX1.
OS  Lithobius forficatus.
OG  Mitochondrion.
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Myriapoda; Chilopoda;
OC  Pleurostigmophora; Lithobiomorpha; Lithobiidae; Lithobius.
OX  NCBI_TaxID=7552;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=20558576; PubMed=11095730;
RA  Lavrov D.V., Brown W.M., Boore J.L.;
RT  "A novel type of RNA editing occurs in the mitochondrial tRNAs of the
RT  centipede Lithobius forficatus.";
RL  Proc. Natl. Acad. Sci. U.S.A. 97:13738-13742(2000).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Lavrov D.V., Brown W.M., Boore J.L.;
RL  Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR  EMBL; AF309492; AAG39986.1; -.
KW  Mitochondrion.
SQ  SEQUENCE  511 AA;  56589 MW;  E01618C6ED6956AA CRC64;

alignment_scores:
         Quality:  76.50            Length:       159
           Ratio:  0.981              Gaps:         9
Percent Similarity:  49.057     Percent Identity:  28.302

alignment_block:
US-09-540-234-1 x Q9G407
```

Align seg 1/1 to: Q9G407 from: 1 to: 511

```
 66 CCTCCAAGCCGGACACGGCTACTGTA...................GC  93
    ||||||||  ||  |||||||                             ||
104 ProProSerLeuThrLeuLeuCysSerAlaAlaValGluSerGlyAl  120

 94 TGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAA 143
    ||||||::  |||  |||   |||  |||  :::|||          ::::||
120 aGlyThrGly...TrpThrValTyrProProLeuSerSerAsnIleSerH 136

144 ACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAA 193
    ::   ::|||||||  |||  :::::::::::   ::::
136 isSer......GlyAlaSerVal.AspMetThrIlePheSerLeuHisLe 150

194 CGCCAATGGCGGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGTGT 243
    |||::::::   ::::|||||||||:::  ::: |||:||
150 uAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleSerThrIleI 167

244 AGCGCTTGATGACTTCC...................TTGCTGTACTCCGCT 275
    :::::  ||||||                         ::  ||||||
167 leAsnMetArgThrSerGlyMetSerPheGluArgValProLeuPheVal 183

276 TGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
    ||||||                     :::|||          |||
184 TrpSerValLysIleThrValIleLeuLeuLeuSer......LeuPr 198

326 TGTACTCCGCTGGCAGCATAGTCTGTTTGCTG...................  356
    ||||||||||||:::||||:::: ||||||
198 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnLeuAsnT 215

357 ...................CCTGTCTGGGGACACCTGTGTACTACAGTACT 389
    |||  |||  |||||||  |||:::  |||:::
215 hrSerPheAspProThrGlyGlyLysAspProIleLeuTyrGlnHis 231

390 TCGTTT...................CCCTTTGTAGTGTGTACTTCTT..... 419
    |||                     |||  |||  :::||||:::||
232 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGl 248

420 ...................AGCCATGTACTC 431
    |||||||:::::
248 yPheGlyMetIleSerHisIleVal 256
```

seq_name: sp_organelle:Q9MQ04

seq_documentation_block:
```
ID  Q9MQ04        PRELIMINARY;        PRT;        265 AA.
AC  Q9MQ04;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE  CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN  COX1.
OS  Tetraselmis aff. maculata.
OG  Mitochondrion.
OC  Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;
OC  Chlorodendrales; Chlorodendraceae; Tetraselmis.
OX  NCBI_TaxID=41887;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BUTCHER;
RX  MEDLINE=99418881; PubMed=10488235;
RA  Burger G., Saint-Louis D., Gray M.W., Lang B.F.;
RT  "Complete sequence of the mitochondrial DNA of the red alga Porphyra
RT  purpurea. Cyanobacterial introns and shared ancestry of red and green
RT  algae.";
RL  Plant Cell 11:1675-1694(1999).
CC  -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC      CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC      3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC      CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC      CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC      AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
```

```
CC      AND COPPER B (BY SIMILARITY).
CC  -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC      FERRICYTOCHROME C (BY SIMILARITY).
CC  -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC  -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR  EMBL; AF116777; AAF43780.1; -.
DR  InterPro; IPR000883; -.
DR  Pfam; PF00115; COX1; 1.
DR  PRINTS; PR01165; CYCOXIDASEI.
KW  Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW  Respiratory chain; Transmembrane.
FT  NON_TER    265    265
SQ  SEQUENCE    265 AA;    28876 MW;    AA0A670488DE9F03 CRC64;
```

alignment_scores:
```
        Quality:       76.00              Length:       145
        Ratio:         1.013              Gaps:           8
Percent Similarity:    51.724    Percent Identity:    26.207
```

alignment_block:
US-09-540-234-1 x Q9MQ04

Align seg 1/1 to: Q9MQ04 from: 1 to: 265

```
 92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    ||||||||:::|| |||   ::: |||  |||          :::::||
126 AlaGlyThrGly...TrpThrValTyrProLeuSerSerIleAlaSe 141

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    |:::                :::|||     :::
141 rHisSer......GlyAlaSerVal.AspLeuAlaIlePheSerLeuHis 155

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGT 241
    :::::::::::    :::::::::::|||||:::::::           |||:::
156 ValSerGlyAlaSerSerIleLeuGlyAlaLeuAsnPheIleThrThrIl 172

242 GTAGCGCTTG.................ATGACTTCCTTGCTGTACTCCG 273
    :::::::                    |||
172 eLeuAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheV 189

274 CTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    ::||||| |||      ||||||:::                 |||:::
189 alTrpSerValLeuIleThrAlaPheLeuLeuLeuSer......Leu 203

324 GCTGTACTCCGCTGGCAGCATAGTCTGTTTGCTG................ 356
    ||||||||||||:::||||:::                :::||||||
204 ProValLeuAlaGlyAlaIleThrMetLeuThrAspArgAsnPheAs 220

357 ...................CCTGGTCTGGGGACACCTGTGTACTACAGTA 387
    |||  |||:::        :::::
220 nThrThrPhePheAspProAlaGlyGlyGlyAspProIle***PheGlnH 237

388 CTTCGTTT...................CCCTTTGTAGTGGTACTTCTT... 419
    |||                        :::||||:::::||
237 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 253

420 ...................AGCCATGTACTCGTGTCT 437
    ||||||||:::::||
254 AlaPheGlyIleValSerHisValIleSerSer 264
```

seq_name: sp_invertebrate:O16955

seq_documentation_block:
```
ID  O16955        PRELIMINARY;        PRT;        349 AA.
AC  O16955;
DT  01-JAN-1998 (TrEMBLrel. 05, Created)
DT  01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT  01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE  COSMID R11G11.
GN  R11G11.13.
```

seq_name: sp_human:Q9P234

seq_documentation_block:
```
ID   Q9P234        PRELIMINARY;          PRT;     638 AA.
AC   Q9P234;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   KIAA1494 PROTEIN (FRAGMENT).
GN   KIAA1494.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20277482; PubMed=10819331;
RA   Nagase T., Kikuno R., Ishikawa K., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human
RT   genes XVII.The complete sequences of 100 new cDNA clones from brain
RT   which code for large proteins in vitro.";
RL   DNA Res. 7:143-150(2000).
DR   EMBL; AB040927; BAA96018.1; -.
DR   InterPro; IPR001452; -.
DR   Pfam; PF00018; SH3; 2.
DR   PRINTS; PR00452; SH3DOMAIN.
DR   PROSITE; PS50002; SH3; 2.
FT   NON_TER       1      1
SQ   SEQUENCE     638 AA;  65516 MW;  01650B629ACD3053 CRC64;
```

alignment_scores:
```
        Quality:      76.00              Length:    113
        Ratio:        1.310              Gaps:        5
Percent Similarity:   51.327   Percent Identity:  29.204
```

alignment_block:
US-09-540-234-1/rev x Q9P234   ..

Align seg 1/1 to: Q9P234   from: 1 to: 638

```
398  GGGAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGGCAGCAAAC  349
     ||:::   ::   |||||||||||||||   ||||||||||   :::::
183  GlySerThrAspGlnIleAlaHisLeuArgProGlnThrArgProSerVa  199

348  AGACTATGCTGCCAGCGGAGTACAGC....AAGGA............  318
     :   |||        ||||||||||         |||||
199  lTyrValAlaIleTyrProTyrThrProArgLysGluAspGluLeuGluL  216

317  .................................GTCTTCGAGTGCTGCACG...  300
     :::::::||| ||||||||:::::         ||
216  euArgLysGlyMetPheLeuValPheGluArgCysGlnAspGlyTyrTrp  232

299  GTAAACAGCACTCAAATGGCTCCAAGC.....GGAGTACAGCAAGGAAG  256
     ::::::||||:||||    |||      :::::||||||
233  PheLysGlyThrSerMetHisThrSerLysIleGlyValPheProGlyAs  249

255  TCATCAAGCGCTACACAGTAAACAGCGAATGGCTCCGATAACATCG  206
     ::::  |||    |||     :::::  ||||||||||:::||||||
249  nTyrValAlaProValThrArgAlaValThrAsnAlaSerGln......  263

205  ACGCCATTGGCGTTGGCTTGTTAGTCTTGTCACAAACTACGGCTTCACCGACGTCT  156
     ||||:::    |||  :::      ||||||||||:::::|||||||
264  ......AlaLysValProMetSerThrAlaGlyGlnThrSer  275

155  CGAGGGGCATGTTTGAGACGACCCTCAAAAGGCCAGGGA  117
     ||||||| ||||: :::  ||||||| |||||||||:::
276  ArgGlyValThrMetValSerProSerThrAlaGlyGly  288
```

seq_name: sp_human:Q9HAM2

seq_documentation_block:

```
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=9415C718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Davidson S., Wohldmann P., Bauer C.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF022976; AAC69087.1; -.
DR   InterPro; IPR000276; -.
SQ   SEQUENCE    349 AA;  39882 MW;  43EC5A1DB130CBDD CRC64;
```

alignment_scores:
```
        Quality:      76.00              Length:    108
        Ratio:        1.070              Gaps:        9
Percent Similarity:   65.741   Percent Identity:  32.407
```

alignment_block:
US-09-540-234-1/rev x O16955   ..

Align seg 1/1 to: O16955   from: 1 to: 349

```
305  TGCACGGTAAACAGCACTCAAATGGCTCCAAGCCGGAGTACAGCAAGGAAG  256
     ||||||:|||::||||||||     |||||    :::||||||||||||||
181  CysThrThrGluSerThrArg...ArgProSerAlaAlaLysSerValLy  196

255  TCATCAAGCGCTACACAGTAAACAGCCACGAATGGCTCCGATAACATCG  206
     : :::::|||| |||||||||||:::::||     :::
196  sLysArgCysLeu..ThrValAspAsnPheGlnPheLeuProSerLeuAla  212

205  ACGCCA...TTGGCGTTGTTAGTCTTGTCACAAACTACGGCTTCACCGACG  159
     ||| |||:::|||:|||    ::::||||||||||::||   :::||
213  LeuProIleLeuThrLeuPheLeuLeuArgGluMetArgIleIleLysGl  229

158  TCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGGCAGGGACTGTTTC  109
     : !||||    :::||||||:::|||   ||||    :::||||||||
229  n.SerArg.........SerThrLeuGlnArgSerGlyValValVal  241

108  ..........CAATATGCCGTGCCAGTCTACACAGTAGCCGTGTCCGGCTT  71
     ||::::||||  ||         |:::||
242  AspAsnGluGluLysTyrGlyLeuPro...ThrLysLeuIleAlaLeuLe  257

70   GGAGGTGTCTTGTCTTTGTCCTTTCCTTCAGATGACCCACTAGGC.........  31
     ||||||| ||||:  :::  ::: ||||||||
257  uThrIleThr..SerPheLeuAlaAspThrProLeuGlyCysIleGlyIy  273

30   ..GCTCCTGCATACATAAGG  13
     ||| |::::: ||:|:::::
273  leAlaLysAlaPheIleGln  279
```

```
ID   Q9HAM2          PRELIMINARY;     PRT;    712 AA.
AC   Q9HAM2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CDNA FLJ11367 FIS, CLONE HEMBA1000303, HIGHLY SIMILAR TO MUS MUSCULUS
DE   PLENTY OF SH3S (POSH) MRNA (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=WHOLE EMBRYO, MAINLY HEAD;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA   Tanase T., Nomura Y., Togiya S., Komai Y., Hara R., Takeuchi K.,
RA   Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA   Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK021429; BAB13822.1; -.
FT   NON_TER        712
SQ   SEQUENCE    712 AA;  74719 MW;  7353D19433IDF8B3 CRC64;

alignment_scores:
             Quality:       76.00             Length:        113
             Ratio:          1.310            Gaps:            5
Percent Similarity:        51.327    Percent Identity:      29.204

alignment_block:
US-09-540-234-1/rev x Q9HAM2       ..

Align seg 1/1  to: Q9HAM2   from: 1  to: 712

398 GGGAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGGCAGCAAAC 349
    |||:::  |||  :::::|||    ||||||||||||   :::|  |||
433 GlySerThrAspGlnIleAlaHisLeuArgProGlnThrArgProSerVa 449

348 AGACTATGCTGCCAGCGAGTACAGC....AAGGAA.............. 318
    |||     |:::|||    |||::     ::::::
449 lTyrValAlaIleTyrProTyrProArgLysGluAspGluLeuGluL 466

317 ...............GTCTTCGAGTGCTGCACG............... 300
                   ||||||| |||
466 euArgLysGlyGluMetPheLeuValPheGluArgCysGlnAspGlyTrp 482

299 GTAAACAGCACTCAAATGGCTCCAAGC.....GGAGTACAGCAAGGAAG 256
    :::|||||::|||:||    |||::       |||::
483 PheLysGlyThrSerMetHisThrSerLysIleGlyValPheProGlyAs 499

255 TCATCAAGCGCTACACAGTAAACAGCACACAGAATGGCTCCGATAACATCG 206
    :::  :||    ::::         ||||||| |||:::||
499 nTyrValAlaProValThrArgAlaValThrArgAsnAlaSerGln...... 513

205 ACGCCATTGGCGTGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTCT 156
         :::|||:|||:|||    ||||||:::::::|||:::    ||||||
514 ..........AlaLysValAlaProMetSerThrAlaGlyValPheProGlyAs
514                AlaLysValProMetSerProGlyGlnThrSer 525

155 CGAGGGGCATGTTTGAGACGACGACCCTCAAAAGGCCAGGGA 117
    ||||||::: :: |||  ||||||  :: |||
526 ArgGlyValThrMetValSerProSerThrAlaGlyGly 538

seq_name: sp_organelle:P92619

seq_documentation_block:
ID   P92619          PRELIMINARY;     PRT;    235 AA.
AC   P92619;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
```

```
GN   COX1.
OS   Chlorella vulgaris.
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellaceae; Chlorella.
OX   NCBI_TaxID=3077;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIES-227;
RA   Watanabe K.I., Inagaki Y., Ehara M., Ohama T.;
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIES-227;
RX   MEDLINE=96269926; PubMed=8662206;
RA   Hayashi-Ishimaru Y., Ohama T., Kawatsu Y., Nakamura K., Osawa S.;
RT   "UAG is a sense codon in several chlorophycean mitochondria.";
RL   Curr. Genet. 30:29-33(1996).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB011523; BAA25069.1; -.
DR   Mendel; 23591; Chlvu;cox1;23591.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER        1          1
FT   NON_TER        235        235
SQ   SEQUENCE    235 AA;  25383 MW;  E046A625140F0BA0 CRC64;

alignment_scores:
             Quality:       75.50             Length:        115
             Ratio:          1.280            Gaps:            5
Percent Similarity:        51.304    Percent Identity:      26.957

alignment_block:
US-09-540-234-1 x P92619           ..

Align seg 1/1  to: P92619   from: 1  to: 235

174 AGTTTGTACAGAGTAACAACAGCCAATGGCGTCGATGTTATCGGAGCCAT 223
    :::::::::::|||  :::|||  |||:::::::::::|||||||
15  AlaIlePheSerLeuHisLeuAlaGlyAlaSerSerIleMetGlyAlaII 31

224 TCGTGTGCGTGTTTACTGTGTAGCGCTTG...............ATGA 255
    :::::::::||||: |||::: ||||||
31  eAsnLeuIleThrThrIlePheAsnMetArgAlaProGlyMetSerMetH 48

256 CTTCCTTGCTGCTACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCGTGCA 305
    ||| ::::|||| ::||    |||:::||
48  isArgLeuProLeuPheValTrpSerValLeuIleThrAlaPheLeuLeu 64

306 GCACTCGAAGACTTCCTTGCTGCTACTCCGCTGCAGCATAGTCTGTTTGCT 355
    |||::||   |||  ||||||||||||::: |||:::||||||
65  IleLeuSer.....LeuProValLeuAlaGlyIleThrMetLeuLe 79

356 G................................CCTGGTCTGGGGA 369
    |||                              |||    |||
79  uThrAspArgAsnPheAsnThrThrPheLeuAspProAlaGlyGlyGlyA 96
```

```
370 CACCTGTGTACTACAGTACTTCGTTT................CCCTTTGTA 404
        |||:::   |||:::   |||                |||  |||
 96 spProIleLeuTyrGlnHisLeuPheTrpPhePheGlyIleValSerProGluVal 112

405 GTGGTACTACTTCTT................AGCCATGTACTC 431
        :::||||:||||                |||||||||||::
113 TyrIleLeuIleLeuProGlyPheGlyIleValSerHisValIle 127
```

seq_name: sp_organelle:Q9TE71

seq_documentation_block:
```
ID   Q9TE71        PRELIMINARY;        PRT;    353 AA.
AC   Q9TE71;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Cylindrotheca closterium.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Bacillariophyta; Bacillariophyceae;
OC   Bacillariophycidae; Bacillariales; Bacillariaceae; Cylindrotheca.
OX   NCBI_TaxID=2856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCMP 339;
RA   Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT   "Phylogenetic analysis of diatom coxI genes and implications of
RT   fluctuating GC content on mitochondrial genetic code evolution.";
RL   Curr. Genet. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB020222; BAA86607.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1       1
FT   NON_TER     353     353
SQ   SEQUENCE   353 AA;   38203 MW;  B8491C4ACD7F0F6A CRC64;
```

alignment_scores:
```
        Quality:     75.50              Length:     119
        Ratio:        1.198             Gaps:         6
Percent Similarity:  52.941   Percent Identity:   26.891
```

alignment_block:
US-09-540-234-1 x Q9TE71

Align seg 1/1 to: Q9TE71 from: 1 to: 353
```
174 AGTTTGTACAAGACTAACAACCGCCAATGGCCTGCGATCGTTATCGGAGCCAT 223
       |:||||| |||||| :||::  |||::::||||||||||||||
133 AlaIlePheSerLeuHisLeuSerGlyAlaAlaSerIleLeuGlyAlaIl 149

224 TCGTGTGCGTGTTTACTGTGTAGCGCTTGATGACTTCCTTGCTG........ 266
       ||||| |:||||: :||:::  |||   |||
149 eAsnPheIleCysThrIleValAsnMetArgThrGluSerLeuProPheH 166
```

```
267 ..................TACTCCCTGCTTGGAGCCATTTGAGTGCGTGTTTACCGTGCA 305
                          ::||||:   ::||| |||| ::||
166 isLysLeuProLeuPheValTrp..................AlaValPheIleThr 178

306 GCACTCGAAGACTTCCTTGCT..................GTACTCGCTGGCCAGCATAGTCTG 349
       ::||||: |||||||||:::               ||||||||:||
179 AlaIleLeuLeuLeuLeuProValLeuAlaGlyAlaIleThrMe 195

350 TTTGCTG..................CCTGGTC 363
       |||||||                   |||
195 tLeuLeuThrAspArgAsnPheAsnThrThrPheAspProAlaGlyG 212

364 TGGGACACCTGTACTACACAGTACTTCGTTT..................CCC 398
       |||  |||||| |||::: |||                      |||
212 lyGlyAspProValLeuTyrGlnHisLeuTrpPhePheGlyHisPro 228

399 TTTGTAGTGGTACTACTTCTT..................AGCCATGTACT 430
       ||| ::||||:||||                      |||||||||
229 GluValTyrIleLeuIleLeuProGlyPheGlyIleIleSerHisIleVa 245

431 CCGTGTCT 437
       :||||||
245 lValSer 247
```

seq_name: sp_plant:O23578

seq_documentation_block:
```
ID   O23578        PRELIMINARY;        PRT;    383 AA.
AC   O23578;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 42.1 KDA PROTEIN.
GN   AT4G17350.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98121113; PubMed=9461215;
RA   Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA   Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA   Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA   Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terryn N.,
RA   Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA   Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kotter P.,
RA   Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA   Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA   Voukelatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA   Hilbert H., Duesterhoft A., Moores T., Jones J.D.G., Eneva T.,
RA   Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
RA   Delseny M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA   Schueller C., Chalwatzis N.;
RT   "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
RT   Arabidopsis thaliana.";
RL   Nature 391:485-488(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z97343; CAB10516.1; -.
DR   EMBL; AL161546; CAB78738.1; -.
DR   Mendel; 26004; Arath:3306;26004.
DR   InterPro; IPR001849; -.
DR   SMART; SM00233; PH; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   383 AA;   42051 MW;  A876779D7D2E121D CRC64;
```

alignment_scores:
```
        Quality:     75.50              Length:     110
```

Percent Similarity: 57.273   Percent Identity: 27.273

Ratio: 1.198   Gaps: 6

alignment_block:
US-09-540-234-1/rev x O23578 ..

Align seg 1/1 to: O23578 from: 1 to: 383

```
402 CAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCC............ 367
        ||||||||:::   :::::||||||  |||   |||
 53 GlnGlnGlnPheCysValGlyGlnAsnSerProProValLeuPh 69

366 .CCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA.... 322
        |||||| ||||||:::  |||   :::  :::::||||||:
 69 eProAspAlaAlaAlaAspProLeuValAlaGlyLysIleMetAsnSerV 86

321 ..GGAAGTCTTCGAGTG.......................CTGCACGGTAA 295
        |||::||||:::::  ||                    |||   |||:::
 86 alGlyThrArgArgAlaGlyThrLeuSerLysTrpPheHisHisAsnLys 102

294 CAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAGTCATCAAGCGC 245
        :::||||||||::||   ::||||||||   ||||||:||||
103 GluHisSerThrSerThrThrThrAsnLeuLysLysAspLysVa 119

244 TACACAGTAAACAGCACACGAATGGCTCCGATAACATCGACGCCATTGGC 195
        :::||||||       |||  :::|||:::::
119 lArgValGluAsnAlaHisVal...........His.SerAlaValSer 131

194 GTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATG 145
        :::||||:::||||:    ||||| |||||  |||:::::::
132 IleAlaAlaLeuAlaAlaGlyLeuAlaSerValThrSerAlaSerAsnCy 148

144 TTTGAGACGACCCTCAAAAGGGCAGGGA 117
        |||||||::::|||
148 s..................LysGlySerGly 152
```

seq_name: sp_organelle:O63273

seq_documentation_block:
ID O63273   PRELIMINARY;   PRT;   324 AA.
AC O63273;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COI.
OS Cataglyphis velox.
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Formicidae; Formicinae; Cataglyphis.
OX NCBI_TaxID=72786;
RN [1]
RP SEQUENCE FROM N.A.
RA Hasegawa E., Tinaut A., Ruano F.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC   FERRICYTOCHROME C (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR EMBL; AB010932; BAA25709.1; -.
DR InterPro; IPR000883; -.
DR Pfam; PF00115; COX1; 1.
KW Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;

KW Respiratory chain; Transmembrane.
FT NON_TER   1   1
FT NON_TER   324   324
SQ SEQUENCE   324 AA;   35970 MW;   15C213B062F553AB CRC64;

alignment_scores:
        Quality: 75.00   Length: 146
        Ratio: 1.027   Gaps: 9
Percent Similarity: 50.000   Percent Identity: 26.712

alignment_block:
US-09-540-234-1 x O63273 ..

Align seg 1/1 to: O63273 from: 1 to: 324

```
 92 GCTGGCACGGCATAATGGAAAACAGTCCTGCCCTTTG...AGGGTCGT 138
        |||||||||| |||  |||      |||  |||
 40 AlaGlyThrGly..TrpThrIleTyrProProLeuAlaSerAsnIlePh 55

139 CTCAAACATGCCCCTCGAGACGTCGGTGAAGCGCTAGTTGTACAGAGACT 188
        ||  |||    ||||||     |||  :::  ||||::
 55 eHisAsnGlyPro........SerVal.AspLeuThrIlePheSerLeu 68

189 AACAACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCGTGTTTAC 238
        :::  |||:::    :::::||:::||||||||||||  ||
 69 HisIleAlaGlyMetSerSerIleLeuGlyAlaIleAsnPheThrSerTh 85

239 TGTGTAGCGCTG..........................ATGACTTCCTTGCGTGCTACT 270
        ||                            ||:::
 85 rIleLeuAsnMetHisHisAsnLeuSerMetAspLysIleProLeuL 102

271 CCGCTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTC 320
        ::||||||  :::|||||||      :::
102 euValTrpSerIleValMetThrAlaIleArgLeuLeuLeuSer...... 116

321 CTTGCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG.............. 356
        ||| ||||||||||||:::::::   ||||||
117 LeuProValLeuAlaGlyLaIleThrMetLeuThrAspArgAsnLe 133

357 ...............CCTGGTCTGGGGACACCTGTACTACA 384
        ::                      |||    :::  :::  |||:::
133 uAsnThrSerPheAspProSerGlyGlyAspProIleLeuTyrG 150

385 GTACTTCGTTT.................CCCTTTGTGGTACTA 413
        ::                      ||| |||     :::||
150 lnHisLeuPheTrpPhePheAlaHisProGluValTyrIleLeuIleLeu 166

414 .............CTTCTTAGCCATGTACTCCGTGTCT 437
        |||::|||||:::::
167 ProGlyPheGlyLeuIleSerHisIleIleMetAsn 178
```

seq_name: sp_organelle:Q9MJ93

seq_documentation_block:
ID Q9MJ93   PRELIMINARY;   PRT;   406 AA.
AC Q9MJ93;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN COI.
OS Jolonica nipponica.
OG Mitochondrion.
OC Eukaryota; Metazoa; Brachiopoda; Rhynchonelliformea; Rhynchonellata;
OC Terebratellidina; Laqueoidea; Laqueidae; Jolonica.
OX NCBI_TaxID=93904;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JOL;
RX MEDLINE=20318514; PubMed=10860643;
RA Saito M., Kojima S., Endo K.;
RT "Mitochondrial COI sequences of brachiopods: genetic code shared with

```
RT      protostomes and limits of utility for phylogenetic reconstruction.";
RL      Mol. Phylogenet. Evol. 15:331-344(2000).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB026509; BAB08012.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     406    406
SQ   SEQUENCE   406 AA;  44182 MW;  47FB7E599CACDF8E CRC64;

alignment_scores:
              Quality:    75.00          Length:    161
              Ratio:       0.987         Gaps:        7
Percent Similarity:   47.205    Percent Identity:   24.224

alignment_block:
US-09-540-234-1 x Q9MJ93

Align seg 1/1  to: Q9MJ93    from: 1  to: 406

107 TGGAAAACAGTCCCTGCCCTTTGAGGTGTCGTCTCAAACATGCCCCTCGA 156
    ||||||||||    |||::::::||||:::  ::||||||||    :::||
 79 TrpLeuLeuProProAlaLeuLeuLeuLeuSerAlaLeuLeuSerAlaAlaMetGl 95

157 GACG.................................................. 160
    |:::
 95 uAlaGlyAlaGlyThrGlyTrpThrIleTyrProProLeuAlaGlyGlyA 112

161 ...........TCGGTGAAGCCGTAGTTGTTACAAGACTAACAA 193
               :||:|| :::  ::||||||||||        ::
112 laAlaHisSerGlyProAlaVal.AspLeuAlaIlePheSerLeuHisLe 128

194 CGCCAATGGCTCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGTGT 243
    ||||||||||:||||||||  :||::||:::|||||||||:||||||:||
128 uAlaGlyAlaSerIleLeuGlyAlaIleAlaAsnPheIleGlySerValV 145

244 AGCGGCTTG..........................ATGACTTCCTTGCTGCTGTACTCCGCT 275
    :||||||||                          ||:||:::|||||||:||||||||||
145 aIleAsnMetLysValLysGlyMetSerMetSerGlnValProLeuPheVal 161

276 TGGAGCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
    |||||| ||| :::|||    |||||||||||||||:||:::
162 TrpSerValLeuThrValIleLeuLeuLeuLeuLeuSer......LeuPr 176

326 TGTACTCCGCTGGCAGCATAGTCTGTTTGCTG.................... 356
    |:||
176 oValLeuAlaGlyAlaIleThrMetLeuLeuMetAspArgAsnPheSerT 193

357 ..........CCTGTTCTCGGGACACCTGTTACTACACAGTACT 389
              ||||::||| |||::||||||||
193 hrSerPhePheAspProGlyGlyGlyAspProValLeuPheGlnHis 209

390 TCGTTT...........CCCTTTGTAGTGGTACTACTTCTT...... 419
    |||||             ||| |||  :::|||||||||:||
210 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGl 226
```

```
420 ...........AGCCATGTACTCGTGTCT 437
             |||||:::::::::::|||
226 yPheGlyIleValSerHisThrValIleSer 236

seq_name: sp_organelle:Q9XP16

seq_documentation_block:
ID   Q9XP16           PRELIMINARY;        PRT;     410 AA.
AC   Q9XP16;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
OS   Microplitis maturus.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita;
OC   Ichneumonoidea; Braconidae; Microgastrinae; Microplitis.
OX   NCBI_TaxID=93004;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mardulyn P., Whitfield J.B.;
RT   "Phylogenetic signal in the COI, 16S, and 28S genes for inferring
RT   relationships among genera of Microgastrinae (Hymenoptera;
RT   Braconidae); evidence of a high diversification rate in this group of
RT   parasitoids.";
RL   Mol. Phylogenet. Evol. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AF102702; AAD40856.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     410    410
SQ   SEQUENCE   410 AA;  46283 MW;  24039383A2F17324 CRC64;

alignment_scores:
              Quality:    75.00          Length:    142
              Ratio:       1.042         Gaps:        8
Percent Similarity:   50.704    Percent Identity:   26.761

alignment_block:
US-09-540-234-1 x Q9XP16

Align seg 1/1  to: Q9XP16    from: 1  to: 410

 95 GGCACGGCCATAATGGAAAACAGTCCCTGCCCTTTGAGGTGTCGTCTCAAA 144
    ||||||::: ||| |||::||  :::::||||||||||||||:::
 27 GlyThrGly...TrpThrValTyrProLeuSerLeuLeuGlyHi 42

145 CATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAAC 194
    :   |||   |||||  |||||||:::::::::  ::
 42 sSerGlyLeu......SerVal.AspMetSerIlePheSerLeuHisLeu 56

195 GCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGTGTA 244
    |||::::||:||||||||| ||||||::||:::|||||||||:|||||
 57 AlaGlyAlaSerIleMetGlyAlaValAsnPheIleThrThrValTy 73
```

```
245 GGCGTTG.............ATGACTTCCTTGCTGTGTACTCCGCTT 276
        :::::           |||       :::::|
 73 rAsnMetArgSerAsnPheMetAsnMetAspLysIleSerLeuPheIleT 90

277 GGAGCCATTTGAGTGCTGTTACCGTGCGCAGCACTCCGAAGACTTCCTTGCT 326
    |||||||||                        |||:::        |||
 90 rpSerValLeuIleThrAlaIleLeuLeuLeuLeuSer.....LeuPro 104

327 GTACTCGCTGGCAGCATAGTCTGTTTGCTG............... 356
    |||||||||:::|:|||::|:| |||||||
105 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnLeuAsnTh 121

357 ...........CCTGGTCTGGGGACACCTGTACTACAGTACTT 390
                   |||      |||:::       |||:::
121 rSerPheAspProSerGlyGlyGlyAspProIleLeuTyrGlnHisL 138

391 CGTTT...........CCCTTTGAGTGGTACTACTTCTT.......... 419
    |||              |||      |||:::|||:::|||
138 euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly 154

420 ...........AGCCATGTACTC 431
                |||||:::|:|
155 PheGlyIleIleSerHisMetIle 162
```

seq_name: sp_organelle:Q9G4C3

seq_documentation_block:
```
ID   Q9G4C3      PRELIMINARY;      PRT;     503 AA.
AC   Q9G4C3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE SUBUNIT 1 (EC 1.9.3.1).
GN   COX1.
OS   Thraustochytrium aureum.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Labyrinthulida; Thraustochytriidae;
OC   Thraustochytrium.
OX   NCBI_TaxID=42467;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Burger G., Lang B.F., Gray M.W.;
RT   "Phylogenetic relationships of stramenopile algae, based on complete
RT   mitochondrial genome sequences.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF288091; AAG23677.1; -.
KW   Oxidoreductase; Mitochondrion.
SQ   SEQUENCE   503 AA;  55493 MW;  B66D4B73B8D822C1 CRC64;
```

alignment_scores:
```
     Quality:     75.00        Length:      149
     Ratio:        1.087       Gaps:          6
Percent Similarity:  46.309    Percent Identity:  24.832
```

alignment_block:
US-09-540-234-1 x Q9G4C3   ..

Align seg 1/1 to: Q9G4C3   from: 1 to: 503

```
107 TGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCCTCGA 156
    |||                 ::||||||||     :::||
110 TrpLeuLeuProProAlaLeuLeuLeuLeuValSerSerSerLeuValGl 126

157 GACG.............................................. 160
    |:::
126 uAlaGlyAlaGlyThrGlyTrpThrValTyrProLeuSerSerValG 143

161 ...........TCGGTGAAGCCGTAGTTTGTACAAGACTAACAA 193
                 |||||  :::::|||:::::::|||::
143 luAlaHisSerGlyProSerVal.AspLeuAlaPheSerLeuHisIl 159
```

```
194 CGCCAATGGCGTGCGATGTTATCGGAGCCATTCGTGTGTGCTGTTTACTGTGT 243
    |||:::::       :::::|||||:::       |||::     |||::
159 eAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrIleP 176

244 AGCGCTTG................ATGACTTCCTTGCTGTACTCCGCT 275
    |||                     |||
176 heAsnMetArgSerProGlyIleThrMetHisArgLeuProLeuPheCys 192

276 TGGAGCCATTTGAGTGCTGTTACCGTGCGCAGCACTCCGAAGACTTCCTTGC 325
    |||:::||                   |||:::         |||
193 TrpAlaValLeuIleThrAlaPheLeuLeuLeuLeuSer......LeuPr 207

326 TGTACTCGCTGGCAGCATAGTCTGTTTGCTG.................... 356
    |||||||||||:::|||||||:::|||||||
207 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnT 224

357 ...........CCTGGTCTGGGGACACCTGTACTACACAGTACT 389
                 |||  |||:::||||||:::       |||:::
224 hrThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrGlnHis 240

390 TCGTTT.................CCCTTTGAGTGGTACTACTTCTT 419
    |||                    |||  |||:::||:::::|||
241 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuLeu 255
```

seq_name: sp_invertebrate:O17047

seq_documentation_block:
```
ID   O17047      PRELIMINARY;      PRT;     707 AA.
AC   O17047;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PROTEIN PHOSPHATASE WITH EF-HANDS.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97471020; PubMed=9326663;
RA   Sherman P.M., Sun H., Macke J.P., Williams J., Smallwood P.M.,
RA   Nathans J.;
RT   "Identification and characterization of a conserved family of protein
RT   serine/threonine phosphatases homologous to Drosophila retinal
RT   degeneration C.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11639-11644(1997).
CC   -!- CATALYTIC ACTIVITY: A PHOSPHOPROTEIN + H(2)O = A PROTEIN +
CC       ORTHOPHOSPHATE (THIS ENZYME IS SERINE/THREONINE SPECIFIC) (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: TO EF-HAND FAMILY.
CC   -!- SIMILARITY: TO SERINE/THREONINE SPECIFIC PROTEIN PHOSPHATASE (OR
CC       PPP) FAMILY.
DR   EMBL; AF023454; AAB82794.1; -.
DR   HSSP; P08129; 1FJM.
DR   InterPro; IPR000048; -.
DR   InterPro; IPR000934; -.
DR   InterPro; IPR002048; -.
DR   Pfam; PF00036; efhand; 3.
DR   Pfam; PF00149; STphosphatase; 1.
DR   Pfam; PF00612; IQ; 1.
DR   PRINTS; PR00114; STPHPHTASE.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_2.
DR   SMART; SM00156; PP2Ac; 1.
KW   Calcium-binding; Hydrolase; Iron; Manganese.
SQ   SEQUENCE   707 AA;  80330 MW;  1943CE218D0A7EF7 CRC64;
```

alignment_scores:
```
     Quality:     74.50        Length:      192
     Ratio:        0.776       Gaps:          9
Percent Similarity:  50.000    Percent Identity:  23.958
```

alignment_block:

US-09-540-234-1/rev x O17047

Align seg 1/1 to: O17047 from: 1 to: 707

```
442 CGAACAGACACGAGTACATGGCTAAGAAGTAGT..............AC 408
        ||  ||    ||||||  ||::::||               ||
  9 ArgGlnAsnProSerThrGluLeuLysLysSerThrArgAlaThrThrTh 25

407 CACTACAAAGGGAAACGAAGTACTGTAGTACACAGTGTCCCCAGACCAG 358
    ||||||:|||              |||::|
 25 rThrThrSerSerGlnArgAsnAsnTyrAsn...............A 37

357 GCAGCAAACAGACTATCCTGCCAGCGAGTACAGCCAAGAAGTCTTCGAGT 308
    :::::::||||:           ::||||::||||||
 37 sProAsnAsnGlnAsnThrSerSerSerGlyAsnLysLysGluSerSer 53

307 GCTGCACGGTAAACACACTCAAAT..GGCTCCAAGCAGCGAGTACAGCAA 261
    :::::||:: :::|||||||||||         |||||||||||||||
 54 SerSer SerLysGlnHisSerSerLysLysSerLysLysSerAsnSerL 70

260 GGAAGTCATCAAGCCGCTACA.............CAGTAAACA........ 232
    ::||::|||   |||  |||              |||  |||
 70 ysLysAsnArgSerProSerProGlnLeuThrIleLysSerAla 86

231 .......GCACACGAATGGCTCCGA.................TA 212
           |||||||||| |||                     ||
 87 IleLeuIleGlnLysTrpTyrArgArgCysGluAlaArgLeuGluAlaAr 103

211 ACATCGACGCCATTGGCGTTGTTAGTCTTGTTACAAACTACGGC....... 169
    |||:::||            |||:|||||||||||||||
103 gArgArgAlaThrTrpGlnIlePheThrAlaLeuGluTyrAlaGlyGluG 120

168 ..............TTCACCGACGTCTCCAGGGGGCATG 145
                   |||::||||||   |||:::|||
120 lnAspGlnLeuLysLeuTyrAlaLysProSerLeuMetAspSerGluG 136

144 TTTGAGACGACCCTCAAAAGGGCA................ 121
    |||    |||  :: |||  :::
137 AlaGluGluAsnGlyLysGlyValGlyValGluAsnGlyArgAsnSerProLe 153

120 ....GGGACTGTTTCCATTATGCCGTGCCAGCTACAGTACCCGTGCTCCG 75
        |||||||||||  ||||||||   ||::||   ::||||||||||
153 uMetSerAlaLeuSerHisTyrAlaLysProSerLeuMetAspSerGluG 170

 74 GCTTGGAGGTGTCTTTGTCCTTTCCTTCAGATGACCCCACTAGGCGCTCCT 25
    ||  |||||||||||:||||||||:  ::::|||||||||:||:: ||||
170 lyGluThrValLysLysMetLeuGluAspThrSerProThrAsnValAsp 186

 24 GCATACATAAGGGCCCGTGGACCC 1
    |||||||||||: ::||||||||
187 IleAspArgAsnTyrLysGlyPro 194
```

seq_name: sp_organelle:O99383

seq_documentation_block:
```
ID   O99383    PRELIMINARY;    PRT;    353 AA.
AC   O99383;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Ectocarpus sp.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Phaeophyceae; Ectocarpales; Ectocarpaceae;
OC   Ectocarpus.
OX   NCBI_TaxID=74372;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HIROSHIMA, JAPAN;
RA   Ehara M., Inagaki Y., Watanabe K.I., Hayashi-Ishimaru Y., Kawai H.,
RA   Ohama T.;
RT   "Phylogenetic analyses of heterokont algae with emphasis on the AUA
```

```
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C.
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY)
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AF037994; AAC94980.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1      1
FT   NON_TER     353    353
SQ   SEQUENCE   353 AA;   38107 MW;   51E80DAB2B3A01A3 CRC64; -
```

alignment_scores:
```
       Quality:    74.00              Length:    143
         Ratio:     1.014               Gaps:      8
Percent Similarity:  51.049      Percent Identity:  26.573
```

alignment_block:
US-09-540-234-1 x O99383

Align seg 1/1 to: O99383 from: 1 to: 353 .

```
 92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    ||||||||| ||  ::  |||   ||    ||:::||||::|||||||::
109 AlaGlyThrGly..TrpThrValTyrProLeuSerGlyIleGlnAl 124

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAAC 191
    ||||||||||||||||   ::: |||  ::|||||||||||::|||||
124 aHisSer.....GlyProSerVal.AspLeuAlaIlePheSerLeuHis 138

192 AACGCCAATGGCCTCGATGTATCGCAGCCCAATTCGTGTGCTGTTTACTGT 241
    ||||||||||||||||||||:::|||||||:||||:||||  :: ||||:
139 LeuSerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleThrIle 155

242 GTAGCGCTTG.....................ATGACTTCCTTGCTGCTACTCCG 273
    :::::||                          |||  :::  |||||||||
155 ePheAsnMetArgAlaProGlyMetThrMetAspArgValProLeuPheV 172

274 CTTGGAGCCATTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTT 323
    |||||||||  |||   |||    :: :::|||||||||||||::::||
172 alTrpSerValIleThrAlaLeuTyrAlaLeuLeuLeuSer......Leu 186

324 GCTGTACTCGCTGGCAGCATAGTCGTTTGCTG........ 356
    |||||||||||||||||:::|||::|  |||||
187 ProValLeuAlaGlyLyGlyIleThrMetLeuThrAspArgAsnLeuAs 203

357 ................CCTGGTCGTCGGGGACACCTGTACTACAGTA 387
                    |||  |||||  ||||||||||||||||||||
203 nThrThrPhePheAspProAlaGlyGlyAspProValLeuTyrGlnH 220

388 CTTCGTTT.......CCCTTTGTAGTGGTACTACTTCTT... 419
    ||||          :::|||||||||||||||||||||
220 lsLeuTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 236

420 .................AGCCATGTACTC 431
                     ||||||::|||
```

237 GlyPheGlyIleIleSerHisIleLeu 245

seq_name: sp_organelle:Q9TE65

seq_documentation_block:
```
ID   Q9TE65          PRELIMINARY;      PRT;      353 AA.
AC   Q9TE65;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COXI.
OS   Thalassionema nitzschioides.
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Bacillariophyta; Fragilariophyceae;
OC   Fragilariophycidae; Thalassionemataceae; Thalassionema.
OX   NCBI_TaxID=33649;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CCAP1084-1;
RA   Ehara M., Inagaki Y., Watanabe K.I., Ohama T.;
RT   "Phylogenetic analysis of diatom coxI genes and implications of
RT   fluctuating GC content on mitochondrial genetic code evolution.";
RL   Curr. Genet. 0:0-0(1999).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB020228; BAA86613.1; -.
DR   InterPro; IPR000883; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1       1
FT   NON_TER     353     353
SQ   SEQUENCE   353 AA;  38187 MW;  EE00F433E808CF8E CRC64;
```

alignment_scores:
```
         Quality:   74.00        Length:     144
         Ratio:      0.987       Gaps:         8
Percent Similarity: 52.083   Percent Identity: 25.694
```

alignment_block:
US-09-540-234-1 x Q9TE65   ..

Align seg 1/1 to: Q9TE65 from: 1 to: 353

```
 95 GGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAA 144
    ||||||::::    ||||    ::: |||          ::::::::||:
110 GlyThrGly...TrpThrValTyrProLeuSerIleIleAlaHi 125

145 CATCCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAC 194
    :   ||||| ::: :::::::::::::                   :::
125 sSer......GlyGlySerVal.AspLeuAlaIlePheSerLeuHisVal 139

195 GCCAATGGCGTCATGTTATCGGAGCCATTCGTGTGCTGTTTACTGTGTA 244
    ||||||:::|||||||:|::::||||:::   ||:::|| ::::::::
140 SerGlyAlaAlaSerIleLeuGlyAlaIleAsnPheIleCysThrIleeh 156

245 GCGCTTGATGACTTCCTTGCTG.........TACTCCGCTT 276
    ::::::  ::::::  |||            :::|
```

156 eAsnMetArgValLysSerLeuSerPheHisLysLeuProLeuPheValT 173

```
277 GGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGCT 326
    |:: ||   |||  :::::|  |||:::::||:::          |||
173 rpAlaIleLeuIleThrAlaVaLLeuLeuLeuSer......LeuPro 187

327 GTACTCGCTGGCAGCATAGTCGTGTTTGCTG................. 356
    |||||||||||:::::||:::|||::: ||||||
188 ValLeuAlaGlyAlaIleThrMetLeuThrAspArgAsnPheAsnTh 204

357 .............CCTGGTCTGGGACACCTGTACTACAGTACTT 390
                 || |||| |||| ||||:::
204 rThrPheAspProAlaGlyGlyGlyAspProValLeuTyrGlnHisL 221

391 CGTTT..........CCCTTTGTAGTGGTACTACTTCTT...... 419
    |||              ||| |||     ::||||::::::
221 euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly 237

420 ...........AGCCATGTACTCGTGTCT 437
               |||||||||:::||||||
238 PheGlyIleValSerHisValIleValSer 247
```

seq_name: sp_organelle:Q9MJ88

seq_documentation_block:
```
ID   Q9MJ88          PRELIMINARY;      PRT;      406 AA.
AC   Q9MJ88;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (FRAGMENT).
GN   COI.
OS   Terebratulina crossei.
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Brachiopoda; Rhynchonelliformea; Rhynchonellata;
OC   Terebratulidina; Cancellothyroidea; Cancellothyrididae; Terebratulina.
OX   NCBI_TaxID=93898;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TCR;
RX   MEDLINE=20318514; PubMed=10860643;
RA   Saito M., Kojima S., Endo K.;
RT   "Mitochondrial COI sequences of brachiopods: genetic code shared with
RT   protostomes and limits of utility for phylogenetic reconstruction.";
RL   Mol. Phylogenet. Evol. 15:331-344(2000).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS OR IGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O + 4
CC       FERRICYTOCHROME C (BY SIMILARITY).
CC   -!- COFACTOR: TWO HEME (A AND A3) GROUPS AND COPPER B (BY SIMILARITY).
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
DR   EMBL; AB026514; BAB08017.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
DR   PROSITE; PS00077; COX1; 1.
KW   Copper; Heme; Inner membrane; Mitochondrion; Oxidoreductase;
KW   Respiratory chain; Transmembrane.
FT   NON_TER       1       1
FT   NON_TER     406     406
SQ   SEQUENCE   406 AA;  44303 MW;  19FBF27F7289E3CB CRC64;
```

alignment_scores:

```
            Quality:   74.00            Length:   145
            Ratio:     0.987            Gaps:     8
Percent Similarity: 51.724   Percent Identity: 26.897

alignment_block:
US-09-540-234-1 x Q9MJ88

Align seg 1/1  to: Q9MJ88  from: 1  to: 406

  92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
     ||||||||||:::    ||    :::    |||    |  ::::||
  98 AlaGlyThrGly...TrpThrValTyrProProLeuAlaSerIleThrSe 113

 142 AAACATGCCCCTCGAGACGTCGGTGAAGCCCTAGTTGTACAAGACTAAC 191
     |:::    ||||||||    ::    ::::::::||    ::
 113 rHis......AlaGlyAlaSerVal.AspLeuAlaIlePheSerLeuHis 127

 192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGT 241
     |||    |::::::::::|||||||:::    ::    |||||
 128 LeuAlaGlyAlaSerSerIleLeuGlyAlaValAsnPheIleThrSerVa 144

 242 GTAGCGCCTGATGACTTCC.............TTGCTGTACTCCG 273
     |||    |||::    :    :::||
 144 lIleAsnLeuArgThrLysGlyLeuThrMetGluArgLeuProLeuLeuV 161

 274 CTTGGAGCCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAGACTTCCTT 323
     :::||||    :::|||    ||||:||    |||:::    |||
 161 alTrpAlaValLysIleThrValIleLeuLeuLeuSer.....Leu 175

 324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG.............. 356
     ||||||||||||:::||:::    |||||
 176 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 192

 357 .................CCTGGTCTGGGGACACCTGTACTACAGTA 387
                       ||    |||    ||||||    ::::::
 192 nThrSerPhePheAspProAlaGlyLyGlyAspProValLeuPheGlnH 209

 388 CTTCGTT...........CCCTTTGTAGTGGTACTACTTCTT...... 419
     |||    :::||||||||::||||    ::    ||
 209 lsLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 225

 420 ..........AGCCATGTACTCGTGTCT 437
               |||||||||||:::||||
 226 GlyPheGlyIleIleSerHisIleValThrSer 236

seq_name: sp_bacteria:Q9PC48

seq_documentation_block:
ID   Q9PC48         PRELIMINARY;    PRT;    447 AA.
AC   Q9PC48;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PROTON GLUTAMATE SYMPORT PROTEIN.
GN   XF1937.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
```

```
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silva W.A. Jr., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004013; AAF84739.1; -.
DR   InterPro; IPR001991; -.
DR   Pfam; PF00375; SDF; 1.
DR   PRINTS; PR00173; EDTRNSPORT.
SQ   SEQUENCE   447 AA;   47787 MW;   061C7BA3F54582F CRC64;

alignment_scores:
            Quality:   74.00            Length:   104
            Ratio:     1.321            Gaps:     3
Percent Similarity: 53.846   Percent Identity: 26.923

alignment_block:
US-09-540-234-1 x Q9PC48

Align seg 1/1  to: Q9PC48  from: 1  to: 447

 113 ACAGTCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCTCGAGACGTC 162
     |||:::|||::    ||||||||||||:::||    |||    ::||
 311 ThrLeuProThrAlaLeuArgValAlaAspGluMetLysLeuProLysGl 327

 163 GGTGAAGCCGT.........AGTTGTACAAGACTAACAACGCCAATGGC 203
     :::    |||||    |||    ::    |||    ::    ||||||
 327 nIle.SerArgPheValLeuThrIleGlyAlaThrAlaAsnGlnAsnGly 343

 204 GTCGATGTTATCGGAGCCATTCGTCGTGCTGTTTACTGTGTAGGCGCTTGAT 253
     ::    :::::    ||||:::    |||:::|||    ::
 344 ThrAlaLeuPheGluGlyValLeuThrValIlePheLeuAlaGlnPhePheGl 360

 254 GACTTCCTTGCTGTACTCCGCTTGGAGCCCATTTGAGTGCTGTTTACCGTG 303
     :::::|||    :::|||    ::    |||    ::    :::
 360 yValAspLeuSerPheSerGlnGlnAlaMetValMetAla.......... 373

 304 CAGCACTCGAAGACTTCCTTGCTGTACTCGCTGGCAGCATAGTCTGTTTG 353
     ||||||||    ::::||||    |||    |||    :::
 374 ..............................ValCysIle 376

 354 CTGCCGCTCGTCGGGACACCTGCGTACTACTACAGTACTACTTCCTTTCGT 403
     |||    ||:::||||||    |::    ::    ||    ||    ||
 377 LeuGlyGlyIleGlyThrAlaGlyValProSerGlySerLeuProValVa 393

 404 AGTGGTACTA 413
     |:::::::::::
 393 lAlaLeuIle 396

seq_name: sp_organelle:Q9G870

seq_documentation_block:
ID   Q9G870         PRELIMINARY;    PRT;    485 AA.
AC   Q9G870;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CYTOCHROME C OXIDASE SUBUNIT 1 (EC 1.9.3.1).
GN   COX1.
```

```
OS   Malawimonas jakobiformis.
OG   Mitochondrion.
OC   Eukaryota; Malawimonadidae; Malawimonas.
OX   NCBI_TaxID=136089;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Burger G., O'Kelly C.J., Gray M.W., Lang B.F.;
RT   "Comparative analysis of mitochondrial genomes of the ancient jakobid
RT   protists.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295546; AAG13707.1; -.
KW   Oxidoreductase; Mitochondrion.
SQ   SEQUENCE   485 AA;  53654 MW;  48DB4A9A145678E8 CRC64;
```

```
alignment_scores:
            Quality:   74.00          Length:      187
              Ratio:    0.851            Gaps:        9
Percent Similarity:   46.524   Percent Identity:   21.925
```

```
alignment_block:
US-09-540-234-1 x Q9G870
```

```
Align seg 1/1  to: Q9G870  from: 1  to: 485
```

```
 26 GGAGCGCCTAGTGGGTCATCTGAGGAAGGACAAAGACACCTCCAAGCC  75
    ||||||||                       |||   :::::::||
 91 GlyAlaPro.......................AspMetAlaSerPr  98

 76 GGACACGGCTACTGTAGCTGGCACGGCATAATGAAAAACAGTCCCTGCCC 125
    ::: :::::::::::::              |||     |||::::|
 98 oArgSerAsnAsnIleSerSer.......TrpLeuLeuProProSerL 112

126 TTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACG............... 160
    ||||| ::::::||::::  :::|||:::
112 euLeuLeuLeuSerSerSerLeuIleGluValGlyAlaGlyThrGly 128

161 ...........................................TC 162
                                               ||
129 TrpThrValTyrProProLeuSerSerIleGlnSerHisSerGlyProSe 145

163 GGTGAAGCCTAGTTTGTACAAGACTAACAACGCCAATGGCGTCGATGTT 212
    |||| :::  :::::::::::::   :::     :::::::::
145 rVal.AspLeuAlaIlePheSerLeuHisLeuSerGlyAlaGlySerIle 161

213 ATCGGAGCCATTCGTGCTGCTGTTTACTGTGTAGCGCTTG.......... 251
    ::||||||||::::: :::  ||::::: ::::::
162 LeuGlyAlaValAsnPheIleThrThrIlePheAsnMetArgAlaProGl 178

252 .......ATGACTTCCTTGCTGTACTCCGCTTGGAGCCATTTGAGTGCTG 294
           |||  |||    :::::::||   |||   ||
178 yLeuThrMetAsnArgLeuProLeuPheValTrpAlaLeuLeuIleThrA 195

295 TTTACCGTGCAGCACTCGAAGACTTCCTTGCTGCTACTCGCTGGCAGCATA 344
    :::::  |||:::  |||  :||  |||:::|||||::|||
195 laPheLeuIleLeuLeuSer.....LeuProValPheAlaGlyValAlaIle 209

345 GTCTGTTTGCTG.......................CC 358
    :::  ||||||
210 ThrMetLeuLeuThrAspArgAsnPheAsnThrThrPheTyrAspProAl 226

359 TGGTCTGGGGACACCTGTGTACTACAGTACTTCGTTT............ 395
    ||| ||| |||  |||||| |||:::      |||
226 aGlyGlyGlyAspProValLeuTyrGlnHisLeuPheTrpPheLeuGlyH 243

396 ..CCCTTGTAGTGGTACTA.................CTTCTTAGCCAT 425
    ||| |||   :::|||              :::::|||||||
243 isProGluValHisIleLeuIleLeuValProGlyPheIleIleSerHis 259

426 GTACTCGTG 434
    |||::::||
260 ValIleVal 262
```

```
OM of: US-09-540-234-1 to: SwissProt_39:*    out_format : pfs
Date: Sep 7, 2001   5:07 PM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09540234/runat_07092001_145856_19530/app_query.fasta_1.507
-DB-SwissProt_39 -QFMT=fastan -SUFFIX=rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE-LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09540234_@CGN1_1_24 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-540-234-1
Query length: 444
Database: SwissProt_39:*
Database sequences: 93435
Database length: 34255486
Search time (sec): 36.670000
```

```
SwissProt_39:SYR_TREPA  +    66.00   107.72   39.79    589  ! 083803 treponema pallidum.
SwissProt_39:ZG20_XENLA -    66.00   106.55   40.33    675  ! P18714 xenopus laevis (af
SwissProt_39:MUCL_RAT   +    66.00   104.71   41.18    837  ! P98089 rattus norvegicus (r
SwissProt_39:CR2_MOUSE  -    66.00   102.98   42.00   1025  ! P19070 mus musculus (mous
SwissProt_39:MA2X_HUMAN +    66.00   102.08   42.43   1139  ! P49641 homo sapiens (hum

seq_name: SwissProt_39:NTPR_RAT

seq_documentation_block:
ID   NTPR_RAT      STANDARD;       PRT;    661 AA.
AC   P28573;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   SODIUM-DEPENDENT PROLINE TRANSPORTER.
GN   SLC6A7.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=92265305; PubMed=1350201;
RA   Fremeau R.T. Jr., Caron M.G., Blakely R.D.;
RT   "Molecular cloning and expression of a high affinity L-proline
RT   transporter expressed in putative glutamatergic pathways of rat
RT   brain.";
RL   Neuron 8:915-926(1992).
CC   -!- FUNCTION: TERMINATES THE ACTION OF PROLINE BY ITS HIGH AFFINITY
CC       SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN SUBPOPULATIONS OF PUTATIVE
CC       GLUTAMATERGIC PATHWAYS OF RAT BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC       FAMILY (SNF).
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M88111; AAA41541.1; -.
DR   InterPro; IPR000175; -.
DR   Pfam; PF00209; SNF; 1.
DR   PRINTS; PR00176; NANEUSMPORT.
DR   PROSITE; PS00610; NA_NEUROTRAN_SYMP_1; 1.
DR   PROSITE; PS00754; NA_NEUROTRAN_SYMP_2; 1.
DR   PROSITE; PS50267; NA_NEUROTRAN_SYMP_3; 1.
KW   Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW   Symport; Amino-acid transport.
FT   DOMAIN         1     45       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      46     66       1 (POTENTIAL).
FT   TRANSMEM      74     93       2 (POTENTIAL).
FT   TRANSMEM     117    137       3 (POTENTIAL).
FT   DOMAIN       138    214       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     215    233       4 (POTENTIAL).
FT   TRANSMEM     242    259       5 (POTENTIAL).
FT   TRANSMEM     295    312       6 (POTENTIAL).
FT   TRANSMEM     324    345       7 (POTENTIAL).
FT   TRANSMEM     378    397       8 (POTENTIAL).
FT   TRANSMEM     425    443       9 (POTENTIAL).
FT   TRANSMEM     459    479       10 (POTENTIAL).
FT   TRANSMEM     500    519       11 (POTENTIAL).
FT   TRANSMEM     538    556       12 (POTENTIAL).
FT   DOMAIN       557    661       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     182    182       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   661 AA;  73684 MW;  8154F26C55A72DA4 CRC64;
```

```
score_list:
Sequence                   Strd  Orig   ZScore  EScore  Len  ! Documentation
SwissProt_39:NTPR_RAT  +   81.50  136.57  0.8763  661  ! P28573 rattus norvegicus (rat)
SwissProt_39:COX1_ALBCO +  79.00  134.00    1.58  509  ! P48887 albinaria coerulea (l
SwissProt_39:LEUK_MOUSE -  77.00  132.32    2.53  395  ! P15702 mus musculus (mouse)
SwissProt_39:NKCR_HUMAN -  77.00  121.12    2.88 1462  ! P30414 homo sapiens (human)
SwissProt_39:NTPR_HUMAN +  75.00  124.39    4.35  636  ! Q99984 homo sapiens (human).
SwissProt_39:AGSR_BOVIN +  74.00  135.79    4.78  134  ! P56413 bos taurus (bovine).
SwissProt_39:2FH2_DROME -  73.50  108.22    7.32 3005  ! P28167 drosophila melanogas
SwissProt_39:METK_BACSU +  73.00  124.51    6.81  400  ! P54419 bacillus subtilis. s
SwissProt_39:COX1_LUMTE +  73.00  122.38    6.97  513  ! Q34941 lumbricus terrestris
SwissProt_39:COX1_PROWI +  72.50  121.40    7.89  514  ! Q05143 prototheca wickerhami
SwissProt_39:COX1_RHILE +  72.50  121.01    7.93  538  ! Q08855 rhizobium leguminosar
SwissProt_39:ESC1_SCHPO +  72.00  122.31    8.74  413  ! Q04635 schizosaccharomyces
SwissProt_39:CCAD_CHICK +  71.50  107.07   11.63 2190  ! 073700 gallus gallus (chick
SwissProt_39:HRX_HUMAN  -  71.50  101.99   12.32 3969  ! Q03164 homo sapiens (human)
SwissProt_39:SPY2_MOUSE +  70.50  121.74   12.33  315  ! Q9qxv8 mus musculus (mouse)
SwissProt_39:MR15_DROME -  70.50  118.23   14.36  424  ! Q9yoi1 drosophila melanogast
SwissProt_39:GCR2_YEAST -  69.50  115.30   16.61  534  ! Q01722 saccharomyces cerevis
SwissProt_39:FIMC_BORPE -  69.00  110.13   19.71  873  ! P33410 bordetella pertussis.
SwissProt_39:M84D_DROME +  68.50  131.01   17.40   68  ! Q01645 drosophila melanogaste
SwissProt_39:TBA1_PELFA +  68.00  113.82   23.67  453  ! Q40831 pelvetia fastigiata.
SwissProt_39:YK25_CAEEL -  68.00  109.07   24.99  789  ! P34332 caenorhabditis elegan
SwissProt_39:COX1_ACACA +  68.00  108.20   25.24  873  ! 037370 acanthamoeba castella
SwissProt_39:COX1_PHYME +  67.00  111.19   30.55  492  ! Q02211 phytophthora megasper
SwissProt_39:COX1_LOCMI +  67.00  110.84   30.67  512  ! Q36421 locusta migratoria (m
SwissProt_39:COX1_RHISA +  67.00  110.84   30.67  512  ! Q99818 rhipicephalus sanguir
SwissProt_39:CIK3_MOUSE -  67.00  110.58   30.76  528  ! P16390 mus musculus (mouse).
SwissProt_39:MBHI_ALCHY +  67.00  109.19   31.25  621  ! P33374 alcaligenes hydrogenc
SwissProt_39:FABB_HAEIN -  66.50  111.87   33.92  406  ! P43710 haemophilus influenza
SwissProt_39:FRA_DROME  -  66.50  108.60   35.21  595  ! P21525 drosophila melanogaste
SwissProt_39:PRIA_BACSU -  66.50  106.01   36.26  805  ! P94461 bacillus subtilis. pr
SwissProt_39:AGSR_HUMAN +  66.00  120.52   34.40  132  ! 000253 homo sapiens (human).
SwissProt_39:GBA5_DICDI -  66.00  112.25   37.80  347  ! P34043 dictyostelium discoic
SwissProt_39:COX1_PELSU +  66.00  108.89   39.27  514  ! 079672 pelomedusa subrufa (a
SwissProt_39:COX1_MARPO +  66.00  108.75   39.33  522  ! P26856 marchantia polymorpha
SwissProt_39:COX1_BETVU +  66.00  108.72   39.34  524  ! P24794 beta vulgaris (sugar
SwissProt_39:COX1_ORYSA +  66.00  108.72   39.34  524  ! P14578 oryza sativa (rice).
SwissProt_39:COX1_ARATH +  66.00  108.67   39.37  527  ! Q07063 arabidopsis thaliana
SwissProt_39:COX1_OENBE +  66.00  108.67   39.37  527  ! P08743 oenothera bertiana (b
SwissProt_39:COX1_MAIZE +  66.00  108.66   39.37  528  ! P08742 zea mays (maize). cyt
SwissProt_39:COX1_SORBI +  66.00  108.62   39.39  530  ! P05502 sorghum bicolor (sorg
```

alignment_scores:
       Quality: 81.50   Length: 105
       Ratio: 1.455   Gaps: 2
Percent Similarity: 53.333   Percent Identity: 24.762

alignment_block:
US-09-540-234-1 x NTPR_RAT

Align seg 1/1 to: NTPR_RAT   from: 1 to: 661

105 AATGGAAAACAGTCCCTGCCCTTTGAGGTCGTCTCAAACATGCCCCTC 154
    :::||||:::::|||||| :: ||||||
174 AspGlyAsnGlyAlaLeuProLeuAsnLeuSerSerThrValSerProSe 190

155 GAGAGCAGTCGGTGAAGCCGTAGTTGTACAAGACTAACAACGCCAATGGCG 204
    |::::: |||||| :::::::::::|||:
190 rGluGluTyrTrpSerArgTyrValLeuHisIleGlnGlySerGlnGlyI 207

205 TCGATGTTATCGGAGCCATTCGTGTCGTGTTTACTGTGTAGCGCTGATG 254
    ||| |||||||| | |||||||
207 leGlyArgProGlyGluIleArg................. 214

255 ACTTCCTTGCTGCTACTCCGCCTGGAGCCATTGAGTGCTGTTTACCGTGC 304
    |||:::: |||::::::
215 ....TrpAsnLeuCysLeuLeuLeuAl 224

305 AGCACTCGAAGACTTCCTTGCTGCTACTCCGCCTGGAGCCAGCATAGTCTGTTTGC 354
    :|| ||||||:::::||| ||| ;
224 aTrpValIleValPheLeuCysIleLeuLeuLysGly..............V 236

355 TGCCTGCTGCTGGGGACCACCTGTGTACTACAGTACTTCGTTTCCCTTTGTA 404
    :: ::: || |||||||||||:::::||||||:::::::
236 aLysSerSerGlyLysValTyrPheThrAlaThrPheProTyrLeu 252

405 GTGGTACTACTTCTT 419
    ::::::||:||:|||
253 IleLeuLeuMetLeu 257

seq_name: SwissProt_39:COX1_ALBCO

seq_documentation_block:
ID   COX1_ALBCO   STANDARD;   PRT;   509 AA.
AC   P48887;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Albinaria coerulea (Land snail).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;
OC   Clausiliidae; Alopiinae; Albinaria.
OX   NCBI_TaxID=42349;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96120351; PubMed=7498775;
RA   Hatzoglou E., Rodakis G.C., Lecanidou R.;
RT   "Complete sequence and gene organization of the mitochondrial genome
RT   of the land snail Albinaria coerulea.";
RL   Genetics 140:1353-1366(1995).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC   CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC   3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC   4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC   INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.

CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X83390; CAA58307.1; -.
DR   HSSP; P00396; 1OCC.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL   57   57   IRON (HEME A) (PROBABLE).
FT   METAL   236   236   COPPER B (PROBABLE).
FT   METAL   240   240   COPPER B (PROBABLE).
FT   METAL   285   285   COPPER B (PROBABLE).
FT   METAL   286   286   COPPER B (PROBABLE).
FT   METAL   371   371   IRON (HEME A3) (PROBABLE).
FT   METAL   373   373   IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   509 AA;   57034 MW;   07236D13431S1A6B CRC64;

alignment_scores:
       Quality: 79.00   Length: 143
       Ratio: 1.097   Gaps: 8
Percent Similarity: 50.350   Percent Identity: 27.972

alignment_block:
US-09-540-234-1 x COX1_ALBCO

Align seg 1/1 to: COX1_ALBCO   from: 1 to: 509

92 GCTGGCACGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
   ||||||||:::: ||| ::: ||| ||||| ||||:: ||||||
118 AlaGlyThrGly..TrpThrValTyrPro.ProLeu......SerSerS 131

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
   |||| :::::||||:::||||||:::::::|||||:
131 erLeuAlaHisSerGlyAlaSerValAspLeuAlaIlePheSerLeuHis 147

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTCGTGTTTACTGT 241
   :::: ||||||||||||:::::::|||||| ::: |||||:
148 LeuAlaGlyMetSerSerIleLeuGlyAlaIleAsnPheIleThrThrIl 164

242 GTAGCGCTTG........................ATGACTTCCTTGCTGCTACTCCG 273
   : :::::|| ||| :: ::
164 ePheAsnMetArgSerProGlyMetThrMetGluArgValSerLeuPheV 181

274 CTTGGAGCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
   |||||||:::||||:::||:::::||||:::|||||||||||
181 alTrpSerIleLeuValThrValPheLeuLeuLeuSer.......Leu 195

324 GCTGTACCTGCTGGCAGCAGATAGTCGTTTGCTG.............. 356
   |||||||||||||||:::::::::: ||||||
196 ProValLeuAlaGlyValAlaIleThrMetLeuThrAspArgAsnPheAs 212

357 ................CCTGGTCGGGGACACCTGTGTACTACAGTA 387
   ::::: ||| ||| |||:::::
212 nThrSerPhePheAspProAlaGlyGlyAspProIleLeuTyrGlnH 229

388 CTTCGTTT................CCCTTTGTAGTGGTACTACTTCTT... 419
   ||| ||| |||::::||:|::|||
229 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 245

420 AGCCATGTACTC 431
   |||||||||:||
246 GlyPheGlyMetIleSerHisIleLeu 254

```
seq_name: SwissProt_39:LEUK_MOUSE

seq_documentation_block:
ID   LEUK_MOUSE     STANDARD;      PRT;    395 AA.
AC   P15702;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   LEUKOSIALIN PRECURSOR (LEUCOCYTE SIALOGLYCOPROTEIN) (SIALOPHORIN)
DE   (CD43 ANTIGEN) (LY 48) (B CELL DIFFERENTIATION ANTIGEN LP-3).
GN   SPN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2J; TISSUE=Liver;
RX   MEDLINE=90269342; PubMed=2347365;
RA   Cyster J.G., Somoza C., Killeen N., Williams A.F.;
RT   "Protein sequence and gene structure for mouse leukosialin (CD43), a
RT   T lymphocyte mucin without introns in the coding sequence.";
RL   Eur. J. Immunol. 20:875-881(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10.P; TISSUE=Liver;
RX   MEDLINE=90370495; PubMed=2144340;
RA   Dorfman K.S., Litaker K.S., Baecher C.M., Frelinger J.G.;
RT   "The nucleotide sequence of Ly 48 (mouse leukosialin, sialophorin):
RT   the mouse homolog of CD43.";
RL   Nucleic Acids Res. 18:4932-4932(1990).
RN   [3]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=94236703; PubMed=7514104;
RA   Shiota J., Nishimura H., Okamoto H., Yu B., Hattori S., Abe M.,
RA   Okada T., Nozawa S., Tsurui H., Hirose S.;
RT   "A unique murine CD43 epitope Lp-3: distinct distribution from
RT   another CD43 epitope S7.";
RL   Cell. Immunol. 155:402-413(1994).
RN   [4]
RP   SEQUENCE OF 345-383 FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=90316596; PubMed=1973410;
RA   Baecher C.M., Dorfman K.S., Mattei M.-G., Frelinger J.G.;
RT   "cDNA cloning and localization of the mouse leukosialin gene (Ly48)
RT   to chromosome 7.";
RL   Immunogenetics 31:307-314(1990).
CC   -!- FUNCTION: ONE OF THE MAJOR GLYCOPROTEINS OF THYMOCYTES AND T
CC       LYMPHOCYTES. PLAYS A ROLE IN THE PHYSICOCHEMICAL PROPERTIES OF
CC       THE T-CELL SURFACE AND IN LECTIN BINDING. PRESENTS CARBOHYDRATE
CC       LIGANDS TO SELECTINS. HAS AN EXTENDED RODLIKE STRUCTURE THAT COULD
CC       PROTRUDE ABOVE THE GLYCOCALYX OF THE CELL AND ALLOW MULTIPLE
CC       GLYCAN CHAINS TO BE ACCESSIBLE FOR BINDING.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: CELL SURFACE OF THYMOCYTES, T LYMPHOCYTES,
CC       NEUTROPHILS, PLASMA CELLS AND MYELOMAS.
CC   -!- PTM: HAS A HIGH CONTENT OF SIALIC ACID AND O-LINKED CARBOHYDRATE
CC       STRUCTURES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use  by  non-profit institutions as long  as its content is in no  way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X17018; CAA34884.1; -.
DR   EMBL; X52609; CAA36840.1; -.
DR   EMBL; S70677; AAB30765.1; -.
DR   EMBL; M30693; AAA39457.1; -.
DR   PIR; S08065; S08065.
DR   PIR; S12702; S12702.
DR   PIR; A43545; A43545.
DR   MGD; MGI:98384; Spn.
KW   Glycoprotein; Transmembrane; Signal; T-cell; Antigen.
FT   SIGNAL        1     19
FT   CHAIN        20    395       LEUKOSIALIN.
FT   DOMAIN       20    248       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    249    271       POTENTIAL.
FT   DOMAIN      272    395       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    167    167       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   395 AA;  40038 MW;  369F201B04DBC055 CRC64;
```

```
alignment_scores:
                    Quality:   77.00            Length:     140
                      Ratio:    1.132             Gaps:       7
        Percent Similarity:   48.571      Percent Identity:  30.714

alignment_block:
US-09-540-234-1/rev x LEUK_MOUSE

Align seg 1/1  to: LEUK_MOUSE     from: 1  to: 395

412 AGTACCACTACAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCCCCAG 363
    |||::||   ||:::|||    ||      ||||||      |||||||
179 SerSerLysThrSerGlyProProValThrAlaThrGlySerLeuGl 195

362 ACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGG......  320
    ||::|||:::|||      |||||||||::||||||
195 yProSerSerGluMetHisGlyLeuProAlaThrThrAlaThrSerSerV 212

319 ..AAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG 272
      ||::||:|||||||::|||||    |||::|||
212 aGluSerSerValAlaArgGlyThrSer.................... 222

271 GAGTACAGCAAGGAAGTCATCAAGCCGCTACACAGTAAACAGCACACGAAT 222
    |||   ::||||   ||::    |||:::|||||||    :::||||||
223 ...........ValSerSerArgLysThrSerThrThrSerThrGl 234

221 GGCTCCGATAACATCG...ACGCCA..................... 200
    :  |||||||||::::  ::||||
234 nAspProIleThrThrArgSerProSerGlnGluSerGlyMetLeuL 251

199 ......................................TTGGCGTTG 191
                                           ::||||||
251 euValProMetLeuIleAlaLeuValValValLeuAlaLeuValAlaLeu 267

190 TTAGTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTG 141
    ||||:::|||   |||   |||   |||   ||||| :::  :::::
268 LeuLeuLeuTrpArgGlnArgGlnLysArgArg.ThrGlyAlaLeu... 282

140 AGACGACCCTC...........AAAAGGGCAGGAGCGTGTTTTCCATTAT 103
    ||||||                |||||||   |||:::|||  ::
283 .....ThrLeuSerGlyGlyGlyGlyLysArgAsnGlyValValAspAlaTrp 297

102 GCCGTGCCAGCTACAGTA 85
    |||   ||||||||||
298 AlaGlyProAlaArgVal 303
```

```
seq_name: SwissProt_39:NKCR_HUMAN

seq_documentation_block:
ID   NKCR_HUMAN     STANDARD;      PRT;   1462 AA.
AC   P30414;
DT   01-APR-1993 (Rel. 25, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   NK-TUMOR RECOGNITION PROTEIN (NATURAL-KILLER CELLS CYCLOPHILIN-
DE   RELATED PROTEIN) (NK-TR PROTEIN).
GN   NKTR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
816 ValGlnAlaThrGlnSerAlaGlnGluLysGlnGlyGlnMetGl 832

249 AGCGCTCACAGTAAACGCACCGAAT 222
    :||||:||||||||||:::
832 uArgThrHisAsnLysGlnGluLysAsn 841
```

seq_name: SwissProt_39:NTPR_HUMAN

seq_documentation_block:
```
ID   NTPR_HUMAN     STANDARD;      PRT;    636 AA.
AC   Q99884;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   SODIUM-DEPENDENT PROLINE TRANSPORTER (FRAGMENT).
GN   SLC6A7 OR PROT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=95379780; PubMed=7651355;
RA   Shafqat S., Velaz-Faircloth M., Henzi V.A., Whitney K.D.,
RA   Yang-Feng T.L., Seldin M.F., Fremeau R.T. Jr.;
RT   "Human brain-specific L-proline transporter: molecular cloning,
RT   functional expression and chromosomal localization of the gene in
RT   human and mouse genomes.";
RL   Mol. Pharmacol. 48:219-229(1995).
CC   -!- FUNCTION: TERMINATES THE ACTION OF PROLINE BY ITS HIGH AFFINITY
CC       SODIUM-DEPENDENT REUPTAKE INTO PRESYNAPTIC TERMINALS.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC       FAMILY (SNF).
```
```
DR   EMBL; S80071; AAB47007.2; -.
DR   InterPro; IPR000175; -.
DR   Pfam; PF00209; SNF; 1.
DR   PROSITE; PS00610; NA_NEUROTRAN_SYMP_1; 1.
DR   PROSITE; PS00754; NA_NEUROTRAN_SYMP_2; 1.
DR   PROSITE; PS50267; NA_NEUROTRAN_SYMP_3; 1.
KW   Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
KW   Symport; Amino-acid transport.
FT   DOMAIN       1     45       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    46     66       1 (POTENTIAL).
FT   TRANSMEM    74     93       2 (POTENTIAL).
FT   TRANSMEM   117    137       3 (POTENTIAL).
FT   DOMAIN     138    214       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   215    233       4 (POTENTIAL).
FT   TRANSMEM   242    259       5 (POTENTIAL).
FT   TRANSMEM   295    312       6 (POTENTIAL).
FT   TRANSMEM   324    345       7 (POTENTIAL).
FT   TRANSMEM   378    397       8 (POTENTIAL).
FT   DOMAIN     425    443       9 (POTENTIAL).
FT   TRANSMEM   459    479       10 (POTENTIAL).
FT   TRANSMEM   500    519       11 (POTENTIAL).
FT   TRANSMEM   538    556       12 (POTENTIAL).
FT   DOMAIN     557   >636       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD   182    182       N-LINKED (GLCNAC...) (POTENTIAL).
FT   NON_TER    636    636
SQ   SEQUENCE   636 AA;   70892 MW;   9A2D1071F59E409B CRC64;
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Blood;
RX   MEDLINE=93133824; PubMed=8421688;
RA   Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA   Ortaldo J.R.;
RT   "A cyclophilin-related protein involved in the function of natural
RT   killer cells.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RN   [2]
RP   REVISIONS.
RA   Anderson S.K.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC       INVOLVED IN THE FUNCTION OF NK CELLS.
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE VIA ITS N-TERMINUS.
CC   -!- SIMILARITY: CONTAINS A CYCLOPHILIN-LIKE PPIASE DOMAIN.
```
```
DR   EMBL; L04288; AAA35734.2; -.
DR   EMBL; AF184110; AAD56402.1; -.
DR   PIR; A47328; A47328.
DR   HSSP; Q27450; 1A58.
DR   MIM; 161565; -.
DR   InterPro; IPR002130; -.
DR   Pfam; PF00160; pro_isomerase; 1.
DR   PRINTS; PR00153; CSAPPISMRASE.
DR   PROSITE; PS00170; CSA_PPIASE_1; 1.
DR   PROSITE; PS50072; CSA_PPIASE_2; 1.
KW   Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.
FT   DOMAIN       1    176       PPIASE, CYCLOPHILIN-TYPE.
FT   DOMAIN     219    240       ARG/LYS-RICH (BASIC).
FT   DOMAIN     421    457       ARG/LYS-RICH (BASIC).
FT   DOMAIN     970   1010       ARG/LYS-RICH (BASIC).
FT   DOMAIN     194    244       ARG/SER-RICH.
FT   DOMAIN     466    574       ARG/SER-RICH.
FT   DOMAIN     664    814       ARG/SER-RICH.
FT   DOMAIN    1311   1348       ARG-SER TANDEM REPEAT-RICH.
SQ   SEQUENCE  1462 AA;  165676 MW;  D98A1147763EF527 CRC64;
```

alignment_scores:
```
         Quality:    77.00             Length:       76
         Ratio:       1.638            Gaps:          3
Percent Similarity: 61.842   Percent Identity: 28.947
```

alignment_block:
US-09-540-234-1/rev x NKCR_HUMAN

Align seg 1/1 to: NKCR_HUMAN from: 1 to: 1462

```
437 AGACACGAGTACATGGCTAAGAAGTAGTAGTACCAC.........TACAAAGG 397
    ::||:|||  :::||||| |||:||            |||||
772 LysHisSerSerSerGluLysThrLeuHisSerLysTyrValLysGl 788

396 GAAA...CGAAGTACTCTAGTACACAGGTGTCCCCAGACCAGGCAGCAAA 350
    |::  ||||||| ||:::||::      :::::||||||:::::
788 yArgAspArgSerSerCysValArgLysTyrSerGluSerArgSerSerL 805

349 CAGACTATGCTGCCAGCGAGTACAGCAAGGAGAGTCTTCGAGTGCTGCCACG 300
    ||||||||:::::||||::                 ||        :::
805 euSerTyrSerSerAspSer...............GluGlnSerSer 815

299 GTAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCA 250
    ||::::||||||| ||| :                :::||||||||:: ::
```

```
alignment_scores:
     Quality:     75.00                Length:     135
       Ratio:      1.210                  Gaps:       3
Percent Similarity: 45.926    Percent Identity: 20.000

alignment_block:
US-09-540-234-1 x NTPR_HUMAN  .

Align seg 1/1  to: NTPR_HUMAN  from: 1  to: 636

 15 TTATGTATGCAGGAGCGCCTAGTGGTCATCTGAAGGAAAGGACAAAGAC  64
    |||||::::::::||:::
165 LeuCysLeuGluHisArgVal.....................  171

 65 ACCTCCAAGCCGGACACCGGCTACTGTAGCTGGCACGGCATAATGGAAAAC  114
                              ::||||::::
172 ........................SerLysAspGlyAsnG  177

115 AGTCCCTGCCCTTTGAGGGTGCGTCTCAAACATCCCCTCGAGACGTCCG  164
    ::||||||||      ::      ||||||:::::
177 lyAlaLeuProLeuAsnLeuThrCysThrValSerProSerGluGluTyr  193

165 TGAAGCCCGTAGTTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTAT  214
    |||||      :::            ::::::||:::
194 TrpSerArgTyrValLeuHisIleGlnGlyGlySerGlnGlyIleGlySerPr  210

215 CGGAGCCATTCGTGTGCTGTTTACTGTGTGTAGCCGCTTGATGACTTCCTTGC  264
    ||| |||||||
210 oGlyGluIleArg.....................  214

265 TGTACTCCGCTTGGAGCCATTTGAGTGCTGTGTTACCGTGCCAGCCACTCGAA  314
    ||::::    |||    :::||
215 ........TrpAsnLeuCysLeuLeuLeuAlaTrpValIle  227

315 GACTTCCTTGCTGTACTCCGCTGGCAGCATAGTCGTTTGTTTGCTGCCTGGTCT  364
    |||||::::::|||  ||
228 ValPheLeuCysIleLeuLysGly..............ValLysSerSe  239

365 GGGGACACCTGTGTACTACAGTACTTCGTTTCCCTTTGTAGTGTACTAC  414
    ||    |||||||::::::||||||:|||||||||:::::||||
239 rGlyLysValValTyrPheThrAlaThrPhePheProTyrLeuIleLeuLeuM  256

415 TTCTT  419
    ::|||
256 etLeu  257

seq_name: SwissProt_39:AGSR_BOVIN

seq_documentation_block:
ID  AGSR_BOVIN    STANDARD;       PRT;   134 AA.
AC  P56413;
DT  15-JUL-1998 (Rel. 36, Created)
DT  15-JUL-1998 (Rel. 36, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  AGOUTI RELATED PROTEIN PRECURSOR.
GN  AGRP OR ART OR AGRT.
OS  Bos taurus (Bovine).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC  Bovidae; Bovinae; Bos.
OX  NCBI_TaxID=9913;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Oulmouden A., Petit J.M., Julien R.;
RL  Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: PLAYS A ROLE IN WEIGHT HOMEOSTASIS. MAY PLAY A ROLE IN
CC      THE REGULATION OF MELANOCORTIN RECEPTORS WITHIN THE HYPOTHALAMUS
CC      AND ADRENAL GLAND, AND THEREFORE IN THE CENTRAL CONTROL OF FEEDING
CC      (BY SIMILARITY).
-!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
-!- SIMILARITY: BELONGS TO THE AGOUTI FAMILY.
CC  -----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
DR  EMBL; AJ002025; CAA05148.1; -.
KW  Signal.
FT  SIGNAL        1     20       POTENTIAL.
FT  CHAIN        21    134       AGOUTI RELATED PROTEIN.
FT  DOMAIN       89    131       CYS-RICH.
FT  DISULFID     89    104       BY SIMILARITY.
FT  DISULFID     96    110       BY SIMILARITY.
FT  DISULFID    103    121       BY SIMILARITY.
FT  DISULFID    107    131       BY SIMILARITY.
FT  DISULFID    112    119       BY SIMILARITY.
SQ  SEQUENCE    134 AA;   14706 MW;  F4B7AE1458B6A24B CRC64;

alignment_scores:
     Quality:     74.00                Length:      85
       Ratio:      2.312                  Gaps:       4
Percent Similarity: 37.647    Percent Identity: 27.059

alignment_block:
US-09-540-234-1 x AGSR_BOVIN  .

Align seg 1/1  to: AGSR_BOVIN  from: 1  to: 134

194 CGCCAATGGCGTCGATGTTATCGGAGCCAT...TCGTGTGCTGTTTACTG  240
    |||::::  ||||||||  |||  |||  ||||||      :::
 84 ArgSerProArgArgCysValArgLeuHisGluSerCysLeuGlyHisGl  100

241 TGTAGCCGCTTGATGACTTCCTTGCTGTACTCCGCTTGGAGCCATTTGAGT  290
    |||
100 nVal.............................ProC  103

291 GCTGTTTACCGTGCCAGCACTCGAAGACTTCCTTGCTGTACTCCGCTGGCAG  340
    ||||| ||||||::
103 ysCysAspProCysAla.........................  108

341 CATAGTCGTTTGTTGCTGCCTGGTCTGGGGACACCTGTGTACTACAGTACTT  390
                                  |||||      :::::||
109 ..................................ThrCysTyrCysArgPhePh  115

391 CGTTTCCCTTTGT...........AGTGGTACTACTTCTTAGCCATGTA  428
    |   :::::||||             |||||||:::::|||||||
115 eAsnAlaPheCysTyrCysArgLysLeuGlyThrThrThrAsnProCyss  132

429 CTCGT 433
    ::|||
132 erArg 133

seq_name: SwissProt_39:ZFH2_DROME

seq_documentation_block:
ID  ZFH2_DROME    STANDARD;       PRT;   3005 AA.
AC  P28167;
DT  01-OCT-1994 (Rel. 30, Created)
DT  01-OCT-1994 (Rel. 30, Last sequence update)
DT  01-OCT-1994 (Rel. 30, Last annotation update)
DE  ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN  ZFH-2.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92001539; PubMed=1680376;
RA  Fortini M.E., Lai Z., Rubin G.M.;
```

RT   "The Drosophila zfh-1 and zfh-2 genes encode novel proteins
RT   containing both zinc-finger and homeodomain motifs.";
RL   Mech. Dev. 34:113-122(1991).
CC   -!- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC   NERVOUS SYSTEM.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.
CC   -!- SIMILARITY: CONTAINS THREE HOMEOBOX DOMAINS.
CC   ---
DR   EMBL; M63450; AAA29051.1; -.
DR   PIR; S27817; S27817.
DR   PIR; S33642; S33642.
DR   HSSP; P15822; 4ZNF.
DR   TRANSFAC; T00920; -.
DR   FlyBase; FBgn0004607; zfh2.
DR   InterPro; IPR000822; -.
DR   InterPro; IPR001356; -.
KW   Zinc-finger; Metal-binding; DNA-binding; Homeobox; Nuclear protein;
KW   Repeat.
FT   ZN_FING   133   156   C2H2-TYPE.
FT   ZN_FING   559   582   C2H2-TYPE.
FT   ZN_FING   614   638   C2H2-TYPE.
FT   ZN_FING   732   756   C2H2-TYPE.
FT   ZN_FING   897   916   C2H2-TYPE (DEGENERATE).
FT   ZN_FING   940   964   C2H2-TYPE.
FT   ZN_FING   999   1023   C2H2-TYPE.
FT   ZN_FING   1074   1098   C2H2-TYPE.
FT   ZN_FING   1210   1233   C2H2-TYPE.
FT   ZN_FING   1341   1365   C2H2-TYPE.
FT   ZN_FING   1438   1462   C2H2-TYPE.
FT   ZN_FING   1477   1500   C2H2-TYPE (DEGENERATE).
FT   ZN_FING   1513   1535   C2H2-TYPE.
FT   ZN_FING   1541   1564   C2H2-TYPE.
FT   DNA_BIND   1797   1856   HOMEOBOX 1.
FT   DNA_BIND   2154   2213   HOMEOBOX 2.
FT   ZN_FING   2234   2256   C2H2-TYPE.
FT   ZN_FING   2371   2393   C2H2-TYPE.
FT   DNA_BIND   2760   2819   HOMEOBOX 3.
SQ   SEQUENCE   3005 AA; 332056 MW; 7E402052BBCE6847 CRC64;

alignment_scores:
        Quality: 73.50   Length: 121
          Ratio: 1.185   Gaps: 3
Percent Similarity: 51.240   Percent Identity: 21.488

alignment_block:
US-09-540-234-1/rev x ZFH2_DROME   ..

Align seg 1/1 to: ZFH2_DROME from: 1 to: 3005

410 TACCACTACAAAGGAAACGAAGTACTGTAGTACACAGG..............372
    :::||||||||| :::::::::| :::|||
566 TrpHisTyrLysTyrGlnGluThrLeuGluIleHisMetArgGluLysHi 582

371 ....................TGTCCCCAGACCAGGCAGCAAA 350
                        :::||||||::
582 sProAspGlyGluSerAlaCysGlyTyrCysLeuAlaGlyGlnHisP 599

349 CAGACTATGTCTGCCAGCCAG..........................TACACAGCAAGGAAGTC 315
    ||| :::||| |||::|
599 roArgLeuAlaArgGlyGluSerTyrSerCysGlyTyrLysProTyrArg 615

314 TTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTACA 265
    ||| |||:::   :::||| ::: :|||:::
616 CysGluIleCysAsnTyrSerThrThrLysGlyAsnLeuSerIleH1 632

264 GCAAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGAATGGCTCCG 215
    :::||| |||:::||| ||||:::|| ::: :|||:::|||::|||
632 sMetGlnSerAspLysHisLeuAsnAsnMetGlnGluLeuAsnSerSerG 649

214 ATAACATCGACGCCATTGGCGTTGTTAGT..........CTTGTA 180
    :||:|| ||| :::   :|||:::||| |||:::
649 lnAsnMetValAlaAlaAlaAlaAlaAlaValThrGlyLysLeuLeu 665

179 CAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACGACCCTC 130
    :::::|||||||||| :::|||:::|||||| ||| :: :|||
666 LeuSerSerSerProGlnValThrAlaAlaCysProSerAsnSerG1 682

129 AAAAGGCAGGGA 117
    :::||| |||
682 yserGlyAlaGly 686

seq_name: SwissProt_39:METK_BACSU

seq_documentation_block:
ID   METK_BACSU   STANDARD;   PRT;   400 AA.
AC   P54419; O34566;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   S-ADENOSYLMETHIONINE SYNTHETASE (EC 2.5.1.6) (METHIONINE
DE   ADENOSYLTRANSFERASE) (ADOMET SYNTHETASE).
GN   METK OR METE.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / PY79;
RX   MEDLINE=96345628; PubMed=8755891;
RA   Yocum R., Perkins J.B., Howitt C.L., Pero J.;
RT   "Cloning and characterization of the metE gene encoding S-
RT   adenosylmethionine synthetase from Bacillus subtilis.";
RL   J. Bacteriol. 178:4604-4610(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98048467; PubMed=9387221;
RA   Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RT   "Sequencing and functional annotation of the Bacillus subtilis genes
RT   in the 200 kb rrnB-dnaB region.";
RL   Microbiology 143:3431-3441(1997).
CC   -!- FUNCTION: CATALYZES THE FORMATION OF S-ADENOSYLMETHIONINE FROM
CC   METHIONINE AND ATP.
CC   -!- CATALYTIC ACTIVITY: ATP + L-METHIONINE + H(2)O = ORTHOPHOSPHATE +
CC   PYROPHOSPHATE + S-ADENOSYL-L-METHIONINE.
CC   -!- PATHWAY: ACTIVATED METHYL CYCLE.
CC   -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE ADOMET SYNTHETASE FAMILY.
CC   ---
DR   EMBL; U52812; AAB17066.1; -.
DR   EMBL; AF008220; AAC00242.1; -.

```
DR   EMBL; Z99119; CAB15033.1; -.
DR   HSSP; P04384; 1XRC.
DR   SubtiList; BG11840; metK.
DR   InterPro; IPR002133; -.
DR   Pfam; PF00438; S-AdoMet_synt; 1.
DR   PROSITE; PS00376; ADOMET_SYNTHETASE_1; 1.
DR   PROSITE; PS00377; ADOMET_SYNTHETASE_2; 1.
KW   Transferase; One-carbon metabolism; ATP-binding.
FT   NP_BIND     129    134       ATP (POTENTIAL).
FT   CONFLICT     26     26       S -> Y (IN REF. 1).
FT   CONFLICT    111    111       Q -> P (IN REF. 1).
FT   CONFLICT    140    140       A -> V (IN REF. 1).
SQ   SEQUENCE   400 AA;  44042 MW;  EF3D97DCD375714A CRC64;

alignment_scores:
        Quality:   73.00         Length:      57
        Ratio:      1.825        Gaps:         0
Percent Similarity:  70.175   Percent Identity:  26.316

alignment_block:
US-09-540-234-1/rev x METK_BACSU

Align seg 1/1  to: METK_BACSU  from: 1  to: 400

365 CAGACCAGGCAGCAAACAGACTATGCTGCCAGCCAGTACGACCAGGAAGT 316
    |||||  ::::::  |||::  ||||:::::||||::  ::::::
 73 GlnThrIleLysGluIleGlyTyrThrArgAlaLysTyrGlyPheAspAla 89

315 CTTCGAGTGCTGCGACGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTAC 266
     ||:::::::|||::::::||:||   :::   ::: :::
 89 aGluThrCysAlaValLeuThrSerIleAspGluGlnSerAlaAspIleA 106

265 AGCAAGGAAGTCATCAAGCGCTACACAGTAAAACAGCACAACGAATGGCTCC 216
    ::||||  |||||||||::::::::::   |||  :::::
106 laMetGlyValAspGlnAlaLeuGluAlaArgGluGlyThrMetSerAsp 122

215 GATAACATCGACGCCATTGGC 195
    ::::::||:::|||||||||
123 GluGluIleGluAlaIleGly 129

seq_name: SwissProt_39:COX1_LUMTE

seq_documentation_block:
ID   COX1_LUMTE    STANDARD;      PRT;     513 AA.
AC   Q34941;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Lumbricus terrestris (Common earthworm).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
OC   Lumbricina; Lumbricidae; Lumbricus.
OX   NCBI_TaxID=6398;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96042914; PubMed=8536978;
RA   Boore J.L., Brown W.M.;
RT   "Complete sequence of the mitochondrial DNA of the annelid worm
RT   Lumbricus terrestris.";
RL   Genetics 141:305-319(1995).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
```

```
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; U24570; AAC46864.1; -.
DR   HSSP; P98002; 1AR1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        58     58       IRON (HEME A) (PROBABLE).
FT   METAL       237    237       COPPER B (PROBABLE).
FT   METAL       241    241       COPPER B (PROBABLE).
FT   METAL       287    287       COPPER B (PROBABLE).
FT   METAL       288    288       COPPER B (PROBABLE).
FT   METAL       373    373       IRON (HEME A3) (PROBABLE).
FT   METAL       375    375       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   513 AA;  56808 MW;  ECB573C300326B70 CRC64;

alignment_scores:
        Quality:   73.00         Length:     159
        Ratio:      1.028        Gaps:         7
Percent Similarity:  44.654   Percent Identity:

alignment_block:
US-09-540-234-1 x COX1_LUMTE

Align seg 1/1  to: COX1_LUMTE  from: 1  to: 513

107 TGGAAAACAGTCCCTGCCCTTTGAGGTGCTCAAACATGCCCCTCGA 156
    |||   ||::||::|||::|||  ::||||||||||| :::||
100 TrpLeuLeuProSerLeuIleLeuLeuValSerSerAlaAlaValGl 116

157 G.............................................. 157
    |
116 uLysGlyAlaGlyThrGlyTrpThrValTyrProProLeuAlaSerAsnL 133

158 ............AGTCGGTGAAGCCGTAGTTTGTACAGACTAACAA 193
                ||||| |||  :::|||||||||:::::  :::
133 euAlaHisAlaGlyProSerVal.AspLeuAlaIlePheSerLeuHisLe 149

194 CGCCAATGGCGCCGATGTTATCGAGCCATTCGTGTGCTGTTTACTGTGT 243
    ||||:::::  :::::|||||||||||||||||::  ||||||
149 uAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrThrValI 166

244 AGCGCTTGATGACTTCCTTGCTG...................TACTCCGCT 275
    ::::: |||   ||| |||                      :::||
166 leAsnMetArgTrpSerGlyLeuArgLeuGluArgIleProLeuPheVal 182

276 TGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
    |||:: ||| |:||:||||:|||  :::|||   |||:::: ||:::||
183 TrpAlaValLeuThrValValLeuLeuLeuLeuSer......LeuPr 197

326 TGTACTCCGTGGCAGCATAGTCTGTTGCTG................... 356
    ||||||||||||| ||:::::||:::
197 oValLeuAlaGlyAlaIleThrMetLeuThrAspArgAsnLeuAsnT 214

357 ...............CCTGGTCTGGGGACACCTGTTACTACAGTACT 389
                   ||| ||| |||::: |||:::::
214 hrSerPheAspProAlaGlyGlyAspProIleLeuTyrGlnHis 230
```

```
alignment_scores:
                   Quality:    72.50         Length:        145
                     Ratio:     0.954          Gaps:           9
Percent Similarity: 52.414    Percent Identity: 26.897

alignment_block:
US-09-540-234-1 x COX1_PROWI

Align seg 1/1    to: COX1_PROWI    from: 1    to: 514

     92 GCTGGCACGGCATAATGAAAACAGTCCTGCCCTTTGAGGGTCGTCTC   141
        ||||||||:::        |||    :::    |||       :::::||
    123 AlaGlyThrGly...TrpThrValTyrProLeuAlaSerIleAlaSe   138

    142 AAACATGCCCCTCGAGACGTCGGTGAAGCGTAGTTGTACAAGACTAAC   191
        |:::            ||||||    :::::||::::::::::    ::
    138 rHisSer......GlyGlySerVal.AspLeuAlaIlePheSerLeuHis 152

    192 AACGCCAATGGCCTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGT   241
        ||||        ||| |||::||||||   |||||   ||||:::||
    153 LeuAlaGlyValSerSerIleLeuGlyAlaIleAsnPheIleCysThrVa   169

    242 GTAGCGCTG........................ATGACTTCCTGCTGGTACTCCG   273
        |   :::|                           ::||||||||| :
    169 lPheAsnMetArgAlaProGlyMetSerMetLeuAspLeuProLeuPhe...V 185

    274 CTTGGAGCCATTGAGTCGTGTTACCGTGCAGCACTCGAAGACTTCCTT   323
        :||||:::    :::::::||||     |||    |||
    185 aTrpAlaValPheIleThrAlaTrpLeuLeuLeuLeu......CysLeu   199

    324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG.............   356
        |||||||||    :|||||::::||||||   |||||
    200 ProValLeuAlaGlyLyGlyIleThrMetLeuLeuThrAspArgAsnPheAs   216

    357 .............CCTGGTCTCGGGACCACCCTGTACTACAGTA   387
        :::|||||||    |||    ||| |:::       |||:::
    216 nThrSerPheAspProAlaGlyLyGlyAspProIleLeuTyrGlnH   233

    388 CTTCGTTT..............CCCTTTGTAGTGGTACTA.........   413
        |||        |||    |||    ::|||
    233 lsLeuTrpPhePheHisProGluValTyrIleLeuIleIlePro   249

    414 ...CTTCTTAGCCATGTACTCCGTGTCT 437
        :::|||||||||:::::||||||||||
    250 GlyPheGlyIleIleSerHisValIleAlaThr 260
```

```
seq_name: SwissProt_39:COX1_PROWI

seq_documentation_block:
ID   COX1_PROWI    STANDARD;    PRT;    514 AA.
AC   Q05143;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1 OR COXI.
OS   Prototheca wickerhamii.
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC   Chlorellaceae; Prototheca.
OX   NCBI_TaxID=3111;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93181269; PubMed=7680126;
RA   Wolff G., Burger G., Lang F.B., Kueck U.;
RT   "Mitochondrial genes in the colourless alga Prototheca wickerhamii
RT   resemble plant genes in their exons but fungal genes in their
RT   introns.";
RL   Nucleic Acids Res. 21:719-726(1993).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X68721; CAA48661.1; -.
DR   PIR; S30291; S30291.
DR   HSSP; P00396; 1OCC.
DR   MendeI; 12219; PROw1;cox1;1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        62     62       IRON (HEME A) (PROBABLE).
FT   METAL       240    240       COPPER B (PROBABLE).
FT   METAL       244    244       COPPER B (PROBABLE).
FT   METAL       289    289       COPPER B (PROBABLE).
FT   METAL       290    290       COPPER B (PROBABLE).
FT   METAL       375    375       IRON (HEME A3) (PROBABLE).
FT   METAL       377    377       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   514 AA;   56874 MW;   C5219A94E8F1BF22 CRC64;
```

```
    390 TCGTTT..............CCCTTTGTAGTGGTACTACTTCTT......   419
        |||        |||    |||::||||:::::||
    231 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGl   247

    420 ..........AGCCATGTACTC   431
                  ||||||||::::::
    247 yPheGlyAlaIleSerHisIleVal   255
```

```
seq_name: SwissProt_39:COX1_RHILE

seq_documentation_block:
ID   COX1_RHILE    STANDARD;    PRT;    538 AA.
AC   Q08855;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME AA3
DE   SUBUNIT 1).
GN   CTAD OR COXA.
OS   Rhizobium leguminosarum.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=384;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94161496; PubMed=8117073;
RA   Gabel C., Bittinger M.A., Maier R.J.;
RT   "Cytochrome aa3 gene regulation in members of the family
RT   Rhizobiaceae: comparison of copper and oxygen effects in
RT   Bradyrhizobium japonicum and Rhizobium tropici.";
RL   Appl. Environ. Microbiol. 60:141-148(1994).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
```

CC   CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC   CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC   AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC   AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC   4 FERRICYTOCHROME C.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 12
CC   POTENTIAL TRANSMEMBRANE DOMAINS.
CC -!- DEVELOPMENTAL STAGE: FREE IN SOIL (NOT AS BACTEROID).
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC ---
CC ---
DR   EMBL; X74341; CAA52388.1; -.
DR   HSSP; P98002; 1AR1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
FT   TRANSMEM   37   57   POTENTIAL.
FT   TRANSMEM   84   104   POTENTIAL.
FT   TRANSMEM   121   141   POTENTIAL.
FT   TRANSMEM   170   190   POTENTIAL.
FT   TRANSMEM   208   228   POTENTIAL.
FT   TRANSMEM   259   279   POTENTIAL.
FT   TRANSMEM   291   311   POTENTIAL.
FT   TRANSMEM   329   349   POTENTIAL.
FT   TRANSMEM   362   382   POTENTIAL.
FT   TRANSMEM   401   421   POTENTIAL.
FT   TRANSMEM   438   458   POTENTIAL.
FT   TRANSMEM   476   496   POTENTIAL.
FT   METAL   82   82   IRON (HEME A) (PROBABLE).
FT   METAL   265   265   COPPER B (PROBABLE).
FT   METAL   269   269   COPPER B (PROBABLE).
FT   METAL   314   314   COPPER B (PROBABLE).
FT   METAL   315   315   COPPER B (PROBABLE).
FT   METAL   400   400   IRON (HEME A3) (PROBABLE).
FT   METAL   402   402   IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   538 AA; 59133 MW; 740C8FE6C902D9BF CRC64;

alignment_scores:
      Quality: 72.50   Length: 115
      Ratio: 1.229   Gaps: 5
Percent Similarity: 51.304   Percent Identity: 25.217

alignment_block:
US-09-540-234-1 to: COX1_RHILE
Align seg 1/1 to: COX1_RHILE   from: 1 to: 538

174 AGTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCAT 223
    ::::::::::   :: |||:::::::: ::::::::|||||||||
171 AlaIlePheSerLeuHisIleAlaGlyAlaSerSerIleLeuGlyAlaIl 187

224 TCGTGTGCTGTTTACCTGTGTAGCGCCTTG............ATGA 255
    ||:: |||::::|||||::::   :::
187 eAsnPheIleThrThrIleLeuAsnMetArgAlaProGlyMetThrLeuH 204

256 CTTCCTTGCTTGCTACTCCGCTTGGAGCCATTGAGTGCTGTTTACCGTGCA 305
    ||||||| ||| ||| :::::::::
204 isLysMetProLeuPheAlaTrpSerValLeuIleThrAlaPheLeuLeu 220

306 GCACTCGAAGACTTCCTTGCTACTCGCTGGCAGCATAGTCTGTTTGCT 355

      |||:::: ||| ||||||||||||::::|:::: ||||
221 LeuLeuSer.....LeuProValLeuAlaGlyAlaIleThrMetLeuLe 235

356 G.................................CCTTGGTCTCGGGGA 369
    | ||| |||
235 uThrAspArgAsnPheGlyThrThrPhePheAlaProGluGlyGlyGlyA 252

370 CACCTGTGTACTACAGTACTTCGTT.............CCCTTTGTA 404
    |||:::: |||:::: ||| ||| |||
252 spProLeuLeuTyrGlnHisLeuPheTrpPhePheGlyHisProGluVal 268

405 GTGGTACTACTTCTT...................AGCCATGTACTC 431
    ::|||::::||| |||||::::::::
269 TyrIleLeuIleLeuProGlyPheGlyMetIleSerHisIleIle 283

seq_name: SwissProt_39:ESC1_SCHPO

seq_documentation_block:
ID   ESC1_SCHPO   STANDARD;   PRT;   413 AA.
AC   Q04635;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ESC1 PROTEIN.
GN   ESC1 OR SPAC56F8.16.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9154314; PubMed=8381348;
RA   Benton B.K., Read M.S., Okayama H.;
RT   "A Schizosaccharomyces pombe gene that promotes sexual
RT   differentiation encodes a helix-loop-helix protein with homology to
RT   MyoD.";
RL   EMBO J. 12:135-143(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Pearson D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: INVOLVED IN THE SEXUAL DIFFERENTIATION PROCESS. MODULATE
CC   THE ABILITY OF THE CELL TO DIFFERENTIATE IN RESPONSE TO THE
CC   NITROGEN STARVATION SIGNAL; IN PARTICULAR IN RESPONSE TO DECREASES
CC   IN THE LEVEL OF CELLULAR CAMP.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC   BHLH PROTEIN.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC   TRANSCRIPTION FACTORS. "MYOGENIC FACTORS" SUBFAMILY.
CC ---
CC ---
DR   EMBL; X69389; CAA49186.1; -.
DR   EMBL; Z69728; CAA93587.1; -.
DR   PIR; S28065; S28066.
DR   HSSP; P25912; 1AN2.
DR   TRANSFAC; T01541; -.
DR   InterPro; IPR001092; -.
DR   Pfam; PF00010; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
KW   Differentiation; Developmental protein; Nuclear protein; DNA-binding.
FT   DNA_BIND   335   345   BASIC DOMAIN.
FT   DOMAIN   346   386   HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ   SEQUENCE   413 AA; 44798 MW; B8BF7DD11545A739 CRC64;

```
alignment_scores:
             Quality:     72.00          Length:     97
             Ratio:       1.180          Gaps:        1
  Percent Similarity:    62.887   Percent Identity:  24.742

alignment_block:
US-09-540-234-1/rev x ESC1_SCHPO ..

Align seg 1/1  to: ESC1_SCHPO  from: 1  to: 413

398 GGGAAACGAAGTACTCTAGTACACAGGTGTCCCAGACCCAGGCAGCAAAC 349
    |||::: |||::: ||| |||::::  ||| |||::::  ||
115 GlySerSerSerTyrValTyrSerValProProThrAsnSerThrTh 131

348 AGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTGCTGCCACGG 299
    ||| :::|||||         ::::||||:::|||
131 rSerGlnAlaSerAlaLysHisSerAlaValProHisArgSerSerGlnP 148

298 TAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCAA 249
    :::||||||  ||||:::|||||  :::||||||
148 heGlnSerThrThrLeuThrProSerThrThrAspSerSerThrAsp 164

248 GCGCTACACAGTAAACAGCACACGAATGGCTCCGATAACATCGACGCCAT 199
    ::: |||         :::|||::::::::::::      :::|||
165 ValSerSerSerAspSerValSerThrSerAlaSerSerSerAsnAlaSe 181

198 TGGCGTGTTAGTCTTGTACAAAACTACGGCTTCACCGACGTCTCGAGGGG 149
    |||:::      |||::::::: |||:::::::::: ||
181 rAsnThrValSerVal....ThrSerProAlaSerSerAlaThrP 196

148 CATGTTTGAGACGACCCTCAAAAGGGCAGGGACTGTTTCC 108
    :::|||||||||::: ||| ||| |||
196 roLeuProAsnGlnProSerGlnGlnGlnPheLeuValSer 209


seq_name: SwissProt_39:CCAD_CHICK

seq_documentation_block:
ID   CCAD_CHICK     STANDARD;      PRT;  2190 AA.
AC   073700; 073701; 073702; 073703; 073704;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   VOLTAGE-DEPENDENT L-TYPE CALCIUM CHANNEL ALPHA-1D SUBUNIT
DE   (CHCACHAID).
GN   CACNAID.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98070847; PubMed=9405708;
RA   Kollmar R., Montgomery L.G., Fak J., Henry L.J., Hudspeth A.J.;
RT   "Predominance of the alpha1D subunit in L-type voltage-gated Ca2+
RT   channels of hair cells in the chicken's cochlea.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14883-14888(1997).
RN   [2]
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=98070848; PubMed=9405709;
RA   Kollmar R., Fak J., Montgomery L.G., Hudspeth A.J.;
RT   "Hair cell-specific splicing of mRNA for the alpha1D subunit of
RT   voltage-gated Ca2+ channels in the chicken's cochlea.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:14889-14893(1997).
CC   -!- FUNCTION: THE ISOFORM ALPHA-1D GIVES RISE TO L-TYPE CALCIUM
CC       CURRENTS. LONG-LASTING (L-TYPE) CALCIUM CHANNELS BELONG TO THE
CC       "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP.
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- ALTERNATIVE PRODUCTS: MANY VARIANTS ARE PRODUCED BY ALTERNATIVE
CC       SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN THE BASILAR PAPILLA OF THE
CC       COCHLEA.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF027602; AAC08304.1; -.
DR   EMBL; AF027603; AAC08305.1; -.
DR   EMBL; AF027604; AAC08306.1; -.
DR   EMBL; AF027605; AAC08307.1; -.
DR   EMBL; AF027606; AAC08308.1; -.
DR   InterPro; IPR000636; -.
DR   InterPro; IPR002077; -.
DR   InterPro; IPR003091; -.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
DR   PRINTS; PR00169; KCHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation; Alternative splicing.
FT   REPEAT        108     404       I.
FT   REPEAT        530     776       II.
FT   REPEAT        894    1176       III.
FT   REPEAT       1213    1496       IV.
FT   DOMAIN        122     121       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      122     140       S1 OF REPEAT I (POTENTIAL).
FT   DOMAIN        141     158       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      159     178       S2 OF REPEAT I (POTENTIAL).
FT   DOMAIN        179     190       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      191     209       S3 OF REPEAT I (POTENTIAL).
FT   DOMAIN        210     230       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      231     249       S4 OF REPEAT I (POTENTIAL).
FT   DOMAIN        250     268       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      269     288       S5 OF REPEAT I (POTENTIAL).
FT   DOMAIN        289     376       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      377     401       S6 OF REPEAT I (POTENTIAL).
FT   DOMAIN        402     544       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      545     564       S1 OF REPEAT II (POTENTIAL).
FT   DOMAIN        565     579       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      580     598       S2 OF REPEAT II (POTENTIAL).
FT   DOMAIN        599     606       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      607     625       S3 OF REPEAT II (POTENTIAL).
FT   DOMAIN        626     635       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      636     654       S4 OF REPEAT II (POTENTIAL).
FT   DOMAIN        655     673       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      674     694       S5 OF REPEAT II (POTENTIAL).
FT   DOMAIN        695     748       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      749     773       S6 OF REPEAT II (POTENTIAL).
FT   DOMAIN        774     907       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      908     926       S1 OF REPEAT III (POTENTIAL).
FT   DOMAIN        927     942       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      943     962       S2 OF REPEAT III (POTENTIAL).
FT   DOMAIN        963     974       CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM   975   993  S3 OF REPEAT III (POTENTIAL).
FT   DOMAIN     994   999  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1000  1019  S4 OF REPEAT III (POTENTIAL).
FT   DOMAIN    1020  1038  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1039  1058  S5 OF REPEAT III (POTENTIAL).
FT   DOMAIN    1059  1148  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1149  1169  S6 OF REPEAT III (POTENTIAL).
FT   DOMAIN    1170  1226  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1227  1245  S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN    1246  1260  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1261  1280  S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN    1281  1297  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1298  1319  S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN    1320  1342  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1343  1362  S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN    1363  1381  CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM  1382  1401  S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN    1402  1468  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM  1469  1493  S6 OF REPEAT IV (POTENTIAL).
FT   DOMAIN    1494  2190  CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       5     8  POLY-GLN.
FT   DOMAIN     674   680  POLY-LEU.
FT   DOMAIN    1154  1160  POLY-ILE.
FT   DOMAIN    1703  1706  POLY-GLU.
FT   DOMAIN     424   441  BINDING TO THE BETA SUBUNIT (BY
FT                         SIMILARITY).
FT   SITE       359   359  CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                         (BY SIMILARITY).
FT   SITE       726   726  CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                         (BY SIMILARITY).
FT   SITE      1122  1122  CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                         (BY SIMILARITY).
FT   SITE      1435  1435  CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                         (BY SIMILARITY).
FT   BINDING   1096  1186  TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING   1449  1515  TO DIHYDROPYRIDINES (BY SIMILARITY).
FT   BINDING   1461  1504  TO PHENYLALKYLAMINES (BY SIMILARITY).
FT   MOD_RES   1504  1504  PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND   1522  1533  POTENTIAL.
FT   CARBOHYD   150   150  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   324   324  N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   459   484  MISSING (IN ISOFORM I-II-LOOP).
FT   VARSPLIC   939   958  ILGYADYVFTSMFTFEIILK -> ILLGYFDYAFTAIFTVEI
FT                         LLK (IN ISOFORM IIIS2).
FT   VARSPLIC  1284  1293  MISSING (IN ISOFORM IVS2-IVS3).
FT   VARSPLIC  1294  1321  GYFSDAWNTFDSLIVIGSIVDVVLSEAD -> HYFTDAWNT
FT                         FDALIVVGSVVDIAITEVN (IN ISOFORM IVS3).
FT   VARSPLIC  1671  1705  AGLRRTLHDIGPEIRRAISCDLQDDEPEENNPDEEE -> VL
FT                         IAHTAQTPFCSPASKLFPFGAEAWLQRAAGVA (IN
FT                         CLONE PSE29/31-2).
FT   VARSPLIC  1706  2190  MISSING (IN CLONE PSE29/31-2).
FT   VARSPLIC  1710  1728  RNGALFGNHINHISSDRRD -> VMSEHGYVIFLLCNMSFI
FT                         E (IN CLONE PSE29/31-1).
FT   VARSPLIC  1729  2190  MISSING (IN CLONE PSE29/31-1).
FT   VARSPLIC  1892  1894  HVY -> NGP (IN CLONE PSE48/154-1).
FT   VARSPLIC  1895  2190  MISSING (IN CLONE PSE48/154-1).
FT   SEQUENCE  2190 AA;  249342 MW;  10680C1CB7708651 CRC64;
```

alignment_scores:
```
Quality:              71.50        Length:             134
Ratio:                 1.254       Gaps:                 5
Percent Similarity:   42.537       Percent Identity:  22.388
```

alignment_block:
```
US-09-540-234-1/rev x CCAD_CHICK

US-09-540-234-1/rev  to: CCAD_CHICK  from: 1  to: 2190

Align seg 1/1  to: CCAD_CHICK  from: 1  to: 2190

376 ACAGGTGTCCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTAC 327
    |||  :::|||  |||||||:::  |||             ||
 22 ThrArgIleProLeuProGlyAspGlyProThr...............Th 33
```

```
326 AGCAAGGAAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCC 277
    :::||||||| ::::: |||:: |||:::|||:::|||
 33 rgInSerAsnSerSerAlaProSerLysGlnThrValLeuSerTrpGlnA 50

276 AAGCG...GAGTACAGCAAGGAGAGTCATCAAGCGCTACAGTAAACAGC 230
    |||        :::::::::   :::::::::|||:::|||:::
 50 laAlaIleAspAlaAlaArgGlnAlaLysAlaAlaGlnAsnMetAsnThr 66

229 ACACGAATGGCTCCGATAACATCGACGCCATTGGCGTTGTTAGTCTTGTA 180
    ||| ||||   |||
 67 ThrThrAlaGlnProValGlySer.......................... 74

179 CAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACGACCCTC 130
    ||| :::|||
 75 ........LeuSerGlnArgLys........................... 79

129 AAAAGGGCAGGGACTGTTTCCATTATGCCGTGCCAGCTACAGTAGCCGT 80
    |||  :::||||||
 80 ..................ArgGlnGlnTyrAlaLysSer 86

79 GTCCGGCTTGGA.........................GGTGTCTTTGT 57
   ::: ||| :::::::::||||
 87 LysLysGlnGlyAsnThrSerAsnSerArgProProArgAlaLeuPheCy 103

56 CCTTTCCTTCAGATGACCCACTAGGCGCTCCTCGCATACATAAGGGCCCGT 7
   |||||||::::::  |||  |||||||:::||||||
103 sLeuSerLeuAsnAsnProIleArgArgAlaCysIleSerLeuValGluT 120

6 GG 5
  ||
120 rp 120
```

seq_name: SwissProt_39:HRX_HUMAN

```
RA   Djaball M., Selleri L., Parry P., Bower M., Young B.D., Evans G.A.;
RT   "A trithorax-like gene is interrupted by chromosome 11q23
RT   translocations in acute leukaemias.";
RL   Nat. Genet. 2:113-118(1992).
RN   [5]
RP   SEQUENCE OF 1251-1538 FROM N.A.
RX   MEDLINE=94215165; PubMed=8162575;
RA   Gu Y., Alder H., Nakamura T., Schichman S.A., Prasad R., Canaani O.,
RA   Saito H., Croce C.M., Canaani E.;
RT   "Sequence analysis of the breakpoint cluster region in the ALL-1 gene
RT   involved in acute leukemia.";
RL   Cancer Res. 54:2326-2330(1994).
RN   [6]
RP   SEQUENCE OF 1251-1654 FROM N.A. (ISOFORM 14P-18B).
RX   MEDLINE=95322025; PubMed=7598802;
RA   Mbangkollo D., Burnett R., McCabe N., Thirman M., Gill H., Yu H.,
RA   Rowley J.D., Diaz M.O.;
RT   "The human MLL gene: nucleotide sequence, homology to the Drosophila
RT   trx zinc-finger domain, and alternative splicing.";
RL   DNA Cell Biol. 14:475-483(1995).
RN   [7]
RP   SEQUENCE OF 1212-1603 FROM N.A.
RX   MEDLINE=95315013; PubMed=7794749;
RA   Marschalek R., Greil J., Lochner K., Nilson I., Siegler G.,
RA   Zweckbronner I., Beck J.D., Fey G.H.;
RT   "Molecular analysis of the chromosomal breakpoint and fusion
RT   transcripts in the acute lymphoblastic SEM cell line with chromosomal
RT   translocation t(4;11).";
RL   Br. J. Haematol. 90:308-320(1995).
RN   [8]
RP   SEQUENCE OF 1421-1540 FROM N.A.
RX   MEDLINE=94020842; PubMed=8414518;
RA   Forster A., Rabbitts T.H.;
RT   "A method for identifying genes within yeast artificial chromosomes:
RT   application to isolation of MLL fusion cDNAs from acute leukaemia
RT   translocations.";
RL   Oncogene 8:3157-3160(1993).
RN   [9]
RP   CHROMOSOMAL TRANSLOCATION WITH GAS7.
RX   MEDLINE=20183971; PubMed=10706619;
RA   Megonigal M.D., Cheung N.-K.V., Rappaport E.F., Nowell P.C.,
RA   Wilson R.B., Jones D.H., Addya K., Leonard D.G.B., Kushner B.H.,
RA   Williams T.M., Lange B.J., Felix C.A.;
RT   "Detection of leukemia-associated MLL-GAS7 translocation early during
RT   chemotherapy with DNA topoisomerase II inhibitors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:2814-2819(2000).
CC   -!- FUNCTION: POSSIBLY ACTS AS A TRANSCRIPTIONAL REGULATORY FACTOR.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: HEART, LUNG, BRAIN AND T AND B LYMPHOCYTES.
CC   -!- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC       T(11;19)(Q23;P13.3) THAT INVOLVES MLL AND MLLT1/ENL;
CC       T(4;11)(Q21;Q23) THAT INVOLVES MLL AND MLLT2/AF4; T(9;11)(P22;Q23)
CC       THAT INVOLVES MLL AND MLLT3/AF9; T(6;11)(Q27;Q23) THAT INVOLVES MLL AND
CC       MLLT4/AF6; T(17;11)(Q23;Q21) THAT INVOLVES MLL AND
CC       MLLT6/AF17; T(X;11)(Q13;Q23) THAT INVOLVES MLL AND MLLT10/AF10;
CC       T(10;11)(P12;Q23) THAT INVOLVES MLL AND AF1Q; T(11;19)(Q23;P13.3)
CC       T(1;11)(Q21;Q23) THAT INVOLVES MLL AND ELL; AND T(11;19)(Q23;P23) THAT INVOLVES MLL
CC       AND GAS7.
CC   -!- SIMILARITY: BELONGS TO THE TRITHORAX FAMILY OF TRANSCRIPTION
CC       FACTORS.
CC   -!- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 PHD ZINC-FINGER DOMAINS.
```

```
DR   EMBL; L04284; AAA58669.1; -.
DR   EMBL; Z69744; CAA93625.1; -.
DR   EMBL; Z69745; CAA93625.1; JOINED.
DR   EMBL; Z69747; CAA93625.1; JOINED.
DR   EMBL; Z69748; CAA93625.1; JOINED.
DR   EMBL; Z69749; CAA93625.1; JOINED.
DR   EMBL; Z69750; CAA93625.1; JOINED.
DR   EMBL; Z69751; CAA93625.1; JOINED.
DR   EMBL; Z69752; CAA93625.1; JOINED.
DR   EMBL; Z69753; CAA93625.1; JOINED.
DR   EMBL; Z69754; CAA93625.1; JOINED.
DR   EMBL; Z69755; CAA93625.1; JOINED.
DR   EMBL; Z69756; CAA93625.1; JOINED.
DR   EMBL; Z69757; CAA93625.1; JOINED.
DR   EMBL; Z69758; CAA93625.1; JOINED.
DR   EMBL; Z69759; CAA93625.1; JOINED.
DR   EMBL; Z69760; CAA93625.1; JOINED.
DR   EMBL; Z69761; CAA93625.1; JOINED.
DR   EMBL; Z69762; CAA93625.1; JOINED.
DR   EMBL; Z69763; CAA93625.1; JOINED.
DR   EMBL; Z69764; CAA93625.1; JOINED.
DR   EMBL; Z69765; CAA93625.1; JOINED.
DR   EMBL; Z69766; CAA93625.1; JOINED.
DR   EMBL; Z69767; CAA93625.1; JOINED.
DR   EMBL; Z69768; CAA93625.1; JOINED.
DR   EMBL; Z69769; CAA93625.1; JOINED.
DR   EMBL; Z69770; CAA93625.1; JOINED.
DR   EMBL; Z69771; CAA93625.1; JOINED.
DR   EMBL; Z69772; CAA93625.1; JOINED.
DR   EMBL; Z69773; CAA93625.1; JOINED.
DR   EMBL; Z69774; CAA93625.1; JOINED.
DR   EMBL; Z69775; CAA93625.1; JOINED.
DR   EMBL; Z69776; CAA93625.1; JOINED.
DR   EMBL; Z69777; CAA93625.1; JOINED.
DR   EMBL; Z69778; CAA93625.1; JOINED.
DR   EMBL; Z69779; CAA93625.1; JOINED.
DR   EMBL; Z69780; CAA93625.1; JOINED.
DR   EMBL; D14540; BAA03407.1; -.
DR   EMBL; L01986; AAA92511.1; -.
DR   EMBL; U04737; AAA18644.1; -.
DR   EMBL; S78570; AAB34770.1; -.
DR   EMBL; X83604; CAA58584.1; -.
DR   EMBL; S66432; AAB28545.1; -.
DR   EMBL; AF231998; AAG26332.2; ALT_TERM.
DR   TRANSFAC; T02337; -.
DR   MIM; 159555; -.
DR   InterPro; IPR001214; -.
DR   InterPro; IPR001487; -.
DR   InterPro; IPR001965; -.
DR   InterPro; IPR002857; -.
DR   Pfam; PF00628; PHD; 3.
DR   Pfam; PF00856; SET; 1.
DR   Pfam; PF02008; zf-CXXC; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS50280; SET; 1.
KW   Proto-oncogene; Chromosomal translocation; DNA-binding; Bromodomain;
KW   Nuclear protein; Zinc-finger; Metal-binding; Transcription regulation;
KW   Alternative splicing.
FT   DOMAIN        17    102       ALA/GLY/SER-RICH.
FT   DNA_BIND     169    180       A.T HOOK (BY SIMILARITY).
FT   DNA_BIND     217    227       A.T HOOK (BY SIMILARITY).
FT   DNA_BIND     301    309       A.T HOOK (BY SIMILARITY).
FT   ZN_FING     1431   1482       PHD-TYPE 1.
FT   ZN_FING     1484   1533       PHD-TYPE 2.
FT   ZN_FING     1566   1627       PHD-TYPE 3.
FT   DOMAIN      1703   1748       BROMODOMAIN (DIVERGENT).
FT   DOMAIN      3840   3969       SET.
FT   DOMAIN       137    143       POLY-GLY.
FT   DOMAIN       561    564       POLY-PRO.
FT   DOMAIN       568    571       POLY-PRO.
FT   SITE        1444   1445       BREAKPOINT FOR TRANSLOCATION TO FORM MLL-
FT                                 GAS7 ONCOGENE.
FT   VARSPLIC    1407   1444       MISSING (IN ISOFORM 14P-18B).
```

```
FT   CONFLICT        144    144       E -> ELTQIPCSWRTKGHIHDKKTEPRLLAWSWCLN
FT                                    (IN REF. 2).
FT   CONFLICT        317    379       GLLINSELEKPQKVRKDKEGTPPLTKEDKTVVRQSPRRIKP
FT                                    VRIIPSSKRTDATIAKQLLQRA -> VSSLILNWKSPRKSG
FT                                    KTRKEHLHLQKKIRQLSDKALEGLSQLGLFLLQKGQMQPLL
FT                                    SNSYRGQ (IN REF. 1).
FT   CONFLICT        556    556       Q -> E (IN REF. 2).
FT   CONFLICT       1487   1487       R -> G (IN REF. 5).
FT   CONFLICT       1603   1603       S -> SGTE (IN REF. 2 AND 7).
FT   CONFLICT       1616   1616       S -> C (IN REF. 6).
FT   CONFLICT       1937   1937       Q -> H (IN REF. 4).
FT   CONFLICT       2181   2181       P -> S (IN REF. 4).
FT   CONFLICT       3718   3718       R -> G (IN REF. 2).
FT   CONFLICT       3759   3759       N -> D (IN REF. 2).
FT   CONFLICT       3813   3813       D -> G (IN REF. 2).
FT   CONFLICT       3901   3901       R -> A (IN REF. 2).

alignment_scores:
       Quality:      71.50        Length:      118
       Ratio:         1.192       Gaps:          6
Percent Similarity:   50.847      Percent Identity:  29.661

alignment_block:
US-09-540-234-1/rev x HRX_HUMAN    ..

Align seg 1/1  to: HRX_HUMAN  from: 1  to: 3969

   367 CCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAA  318
       |||::::|||  |||:::::||||||      |||::: |||||||||||
  3407 ProGlnLeuGlyThrSerGlnThrProSerThrAlaAlaIleThrAlaAl  3423

   317 GTCTTCGAGTGCTGCACGGTAAACACACTCAA................... 285
       ||||||| |||:::::: |||||||||
  3423 aSerSer..IleCysValLeuProSerThrGlnThrThrGlyIleThrAl  3439

   284 .ATGGCTCCAAGCCGA..............................GTAC  266
       ::|||||||||                                      |||
  3439 aAlaSerProSerGlyGluAlaAspGluHisTyrGlnLeuGlnHisValA  3456

   265 AGCAA.....GGAAGTCATCAAGCGCTACACAGTAAACAGCAC........ 228
       ::|||      :::||| :::::||||||||:::|||::
  3456 snGlnLeuLeuAlaSerLysThrGlyIleHisSerSerGlnArgAspLeu  3472

   227 ...............ACGAATGGCTCCGATAAACATCGACGC........ 202
       :::|| ||||||:::::|||||
  3473 AspSerAlaSerGlyProGlnValSerAsnPheThrGlnThrValAspAl  3489

   201 CATTGGCGGTTGTTAGTCTTGTACAA.....................ACTA 173
       |    :::::||| |||
  3489 aProAsnSerMetGlyLeuGlyGlnAsnLysAlaLeuSerSerAlaValG  3506

   172 CGGCTTCACCGACGTCTCGAGGGCATGTTGAGACGACCCTCAAAAGGG   123
       |||||||||||||  |||:::           |||||||:::|||
  3506 lnAlaSerProThrSerProGlyGlySerProSerProSerSerGly  3522

   122 CAG  120
       |||
  3523 Gln  3523
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99429807; PubMed=10498682;
RA   Minowada G., Jarvis L.A., Chi C.L., Neubueser A., Sun X., Hacohen N.,
RA   Krasnow M.A., Martin G.R.;
RT   "Vertebrate sprouty genes are induced by FGF signaling and can cause
RT   chondrodysplasia when overexpressed.";
RL   Development 126:4465-4475(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SWISS WEBSTER;
RX   MEDLINE=9175483; PubMed=10074434;
RA   Tefft J.D., Lee M., Smith S., Leinwand M., Zhao J., Bringas P. Jr.,
RA   Crowe D.L., Warburton D.;
RT   "Conserved function of mSpry-2, a murine homolog of Drosophila
RT   sprouty, which negatively modulates respiratory organogenesis.";
RL   Curr. Biol. 9:219-222(1999).
CC   -!- FUNCTION: MAY FUNCTION AS AN ANTAGONIST OF FIBROBLAST GROWTH
CC       FACTOR (FGF) PATHWAYS AND MAY NEGATIVELY MODULATE RESPIRATORY
CC       ORGANOGENESIS.
CC   -!- SUBCELLULAR LOCATION: ASSOCIATED WITH MICROTUBULES IN UNSTIMULATED
CC       CELLS BUT IS TRANSLOCATED TO THE MEMBRANE RUFFLES IN CELLS
CC       STIMULATED ITH EGF (EPIDERMAL GROWTH FACTOR).
CC   -!- TISSUE SPECIFICITY: IN ADULT, HIGHLY EXPRESSED IN BRAIN, LUNG,
CC       HEART AND AT LOWER LEVELS IN SKELETAL MUSCLE AND KIDNEY. IN
CC       EMBRYO, HIGHLY EXPRESSED IN LUNG EPITHELIAL CELLS, PRIMARILY IN
CC       THE DISTAL AIRWAYS.
CC   -!- DEVELOPMENTAL STAGE: AT E8.5 EXPRESSED IN THE PRIMITIVE STREAK,
CC       ROSTRAL FOREBRAIN, CELLS LATERAL TO THE POSTERIOR HINDBRAIN,
CC       ANTERIOR HINDBRAIN AND DEVELOPING MIDBRAIN. AT E9.5 CONTINUES TO
CC       BE EXPRESSED IN THE ROSTRAL FOREBRAIN AND PRIMITIVE STREAK, AND IS
CC       ALSO DETECTED IN THE BRANCHIAL ARCHES AND THE FORELIMB BUD. AT
CC       E10.5 EXPRESSED IN THE SOMITES, FRONTONASAL PROCESSES, TAILBUD,
CC       AND HINDLIMB BUD.
CC   -!- INDUCTION: BY FGF SIGNALING.
CC   -!- DOMAIN: THE CYS-RICH DOMAIN IS RESPONSIBLE FOR THE LOCALIZATION OF
CC       THE PROTEIN TO THE MEMBRANE RUFFLES.
CC   -!- SIMILARITY: BELONGS TO THE SPROUTY FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF176905; AAD56006.1; -.
DR   EMBL; AF153084; AAD34167.1; -.
DR   MGD; MGI:1345138; Spry2.
KW   Developmental protein; Membrane; Microtubules.
FT   DOMAIN         124    130       POLY-SER.
FT   DOMAIN         178    301       CYS-RICH.
FT   CONFLICT       163    163       V -> I (IN REF. 2).
SQ   SEQUENCE   315 AA;  34623 MW;  81514698EAD809A7 CRC64;
```

```
seq_name: SwissProt_39:SPY2_MOUSE

seq_documentation_block:
ID   SPY2_MOUSE     STANDARD;     PRT;    315 AA.
AC   Q9QXV8; Q9WUQ9;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   SPROUTY HOMOLOG 2 (SPRY-2).
GN   SPRY2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
```

```
alignment_scores:
       Quality:      70.50        Length:      184
       Ratio:         0.881       Gaps:         14
Percent Similarity:   43.478      Percent Identity:  26.087

alignment_block:
US-09-540-234-1 x SPY2_MOUSE    ..

Align seg 1/1  to: SPY2_MOUSE  from: 1  to: 315

     2 GGTCCACGGGCCCTTATGTATGCAGGAGCGCCTAGTGGTCATCTGAAGG   51
       ||| |||:::    |||::::::|| |||||||||||||:||||||:|||
   116 GlySerArgSerThrArgThrSerSerSerGlu.G  132

    52 AAAGGACAAAGACACCTCCAAGCCGGACACGGCTACTGTTAGCTGGCACGG  101
       :::|||  ||| |||  :::||||||    :::|||||||  :::||:||||  ||
```

```
132 lnArgLeuLeuGlyPro.SerPheSerHisGlyProAlaAlaAlaAlaAspGl 148

102 CATAATGGAAAACAGTCCCTGCCCTTTTGAGGTCGTCTCAAA..........  144
    |||||::::: :::|||            :::::::::||||||
148 yIleIleArgValGlnPro........LysSerGluLeuLysProGlyA 162

145 .............................CATGCCCCTCGA.......  156
                                 ||||| |||
162 spValLysProLeuSerLysAspAspLeuGlyLeuHisAlaTyrArgCys 178

157 ...GACGTCGGTGAAGCCGTAGTTGTACAAGACTAACACGCCAATGGC  203
    ||| |||::::: :::   ||| :::||| ||| |||
179 GluAspCysGlyLysCys...LysCysLysGluCysThrTyrProArgPr 194

204 GTCGATGTTATCGGAGGCCATTCGTGGC.............TGTTACTGTGTA 244
    |||  :::::::||||:::||                     |||||||| |||
194 o......LeuProSerAspTrpIleCysAspLysAspLysGlnCysLeu..CysSer 208

245 GCGCCTGATGACTTCCTTGCTGTACTCCGCCTTGGAGCCATTGAGTGCTG  294
                                 ||||
209 AlaGlnAsn..........ValIleAspTyrArgGlyThrCysValCysCy 221

295 TTTA...................................  298
    |:::
221 sValLysGlyLeuPheTyrHisCysSerAsnAspAspGluAspAsnCysA 238

299 ....CCGTGCAGCACTCGAAGACTTCCTTGCTGTACTCGCTGGCAG 340
    |||||||||          ||:: ||||||||||||||||||
238 laAspAsnProCysSerCysSerGlnSerHisCysCysThrArgTrpSer 254

341 ..........CATGTCTGTTGCTCGCTGGTCTGGGACACCTGTGT 378
    ||||||||              ||:::
255 AlaMetGlyValMetSerLeuPheLeu..............ProCysLe 266

379 ACTACAGTACTTC............GTTTCCCTTTGTAGTGGT 409
    :    |||||||              :::::::||||||
266 uTrpCysTyrLeuProAlaLysGlyCysLeuLysLysCysGlnGly 281
```

seq_name: SwissProt_39:MR15_DROME

seq_documentation_block:
ID   MR15_DROME   STANDARD;   PRT;   424 AA.
AC   Q9YOYI; Q9VF99;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   MRG15 PROTEIN.
GN   MRG15 OR CG6363.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bertram M.J., Pereira-Smith O.M.;
RT   "Conservation of the MORF4 related gene family: identification of a
RT   new chromo domain subfamily and novel protein motif.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE MRG FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AF152245; AAD38047.1; -.
DR   EMBL; AE003708; AAF55161.1; -.
DR   FlyBase; FBgn0027378; MRG15.
KW   Nuclear protein.
SQ   SEQUENCE   424 AA; 47194 MW; B0E1F61525208EDD CR C64;

alignment_scores:
                    Quality: 70.00   Length: 67
                       Ratio: 1.842   Gaps: 1
    Percent Similarity: 56.716   Percent Identity: 34.328

alignment_block:
US-09-540-234-1/rev x MR15_DROME

Align seg 1/1 to: MR15_DROME   from: 1 to: 424

```
417 GAAGTAGTACCACTACAAAGGGAAACGAAGAGTACTGTAGTACACAGGTCTC 368
    ||| |||||| :::|||   ||||||   |||||  :::  ::|||||||
 66 GluTrpValProGluAsnArgValLeuLysTyrAsnAspAspAsnValLy  82

367 CCCAGACCAGGCAGCAGCAAACAGACTATGCTGCCAGCGAGTA.......... 328
    :   |||  |||||||  ||     ::||| |||  :::|||
 82 sArgArgGlnGluLeuAlaArgGlnCysGlyGlyArgSerLysAspA  99

327 ..CAGCAAGGAAGTCTTCGAGTGCTGCACGGTAAACAGCCACTCAAATGGC 280
    :::::||||| :::    |||  ::||||||   |||
 99 snLysLysGlySerAlaLysMetGluGlnMetArgAsnGlu 115

279 TCCAAGCCGGAGTACAGCAAGGAAGTCATCAAGCCGCTACACAGTAAACAGC 230
    ||||::  |||||   :::|||:::|||:::|||  |||||
116 SerArgAlaSerThrProSerLysAspSerAsnThrSerGlnSerThrAl 132

229 A 229
```

```
132 a 132

seq_name: SwissProt_39:GCR2_YEAST

seq_documentation_block:
ID   GCR2_YEAST     STANDARD;      PRT;    534 AA.
AC   Q01722;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   GLYCOLYTIC GENES TRANSCRIPTIONAL ACTIVATOR GCR2.
GN   GCR2 OR YNL199C OR N1374.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93375051; PubMed=1508187;
RA   Uemura H., Jigami Y.;
RT   "Role of GCR2 in transcriptional activation of yeast glycolytic
RT   genes.";
RL   Mol. Cell. Biol. 12:3834-3842(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=9524839; PubMed=7725799;
RA   Jonniaux J.-L., Coster F., Purnelle B., Goffeau A.;
RT   "A 21.7 kb DNA segment on the left arm of yeast chromosome XIV
RT   carries WHI3, GCR2, SPX18, SPX19, an homologue to the heat shock gene
RT   SSB1 and 8 new open reading frames of unknown function.";
RL   Yeast 10:1639-1645(1994).
CC   -!- FUNCTION: REQUIRED FOR THE EXPRESSION OF MOST GLYCOLYTIC GENES.
CC       GCR1 AND GCR2 PROBABLY FUNCTION AS A TRANSCRIPTIONAL ACTIVATION
CC       COMPLEX, GCR1 PROVIDING THE SPECIFIC DNA-BINDING FUNCTION AND
CC       GCR2 PROVIDING THE ACTIVATION FUNCTION.
CC   -!- SUBUNIT: FORMS A GCR1/GCR2 COMPLEX.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC   -!- SIMILARITY: SOME, TO GCR1.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D10104; BAA00985.1; -.
DR   EMBL; X78898; CAA55509.1; -.
DR   EMBL; Z71475; CAA96097.1; -.
DR   PIR; S31300; S31300.
DR   SGD; S0005143; GCR2.
KW   Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN      281    288       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN      255    283       ASN-RICH.
SQ   SEQUENCE   534 AA;  58062 MW;  958D4A9393255B1B CRC64;

alignment_scores:
       Quality:     69.50       Length:   126
       Ratio:        0.993      Gaps:       6
Percent Similarity: 55.556      Percent Identity:  23.810

alignment_block:
US-09-540-234-1/rev x GCR2_YEAST    .
Align seg 1/1  to: GCR2_YEAST from: 1  to: 534

433 ACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGAAGTACT 384
     |||||||||:::|::|  ||::::::::::::::::::::
 75 ThrSerThrGlySerAsnAlaLeuThrLysThrAsnSerAspSerAla.. 90
```

```
383 GTAGTACACAGGTGTCCCAGACCAGGCCAGGCAAACAGACTATGCTG.... 338
     :::||||:::|||::  :::|||:::|||::|||
 91 .......AsnGlyThrProAsnGlyAsnSerSerThrSerAlaIleSer 105

337 ..........CCAGGCAGTACAGCAAGGAAGTCTTCGAGTGCTGCA 302
               ||||||:|::|||||||:::||:|||||||:::|||
105 erAsnAlaSerAsnProAlaThrThrGlyAsnAlaSerSerSerAla 121

301 CGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAG....... 260
     :::::|:  ::  |||||:|||::||||||::
122 ThrSerAsnGlyIle...TyrThrGlnAlaGlnTyrSerGlnLeuPheAl 137

259 .GAAGTCATCAAGCGCTAC.....ACAGTAAACAGCACACGAATGGCTC 217
     :::::|  |||  ||| |||  |||::::::::::::||
137 aLysIleSerLysLeuTyrAsnAlaThrLeuSerSerGlySerIleAspA 154

216 CGATAACATCGACGCCATTGGCGTTGCGTTGTTAGTCTTGTACAAA..... 176
     ::::::::::||  |||||||||:::  |||||:|
154 spArgSerThrSerProLysSerAlaIleGluLeuTyrGlnArgPheGln 170

175 ................................................CTACG 171
                                                     :::
171 GlnMetIleLysGluLeuGlnLeuSerPheAspAlaSerProTyrAlaLy 187

170 GCTTCACCGACGTCTCGAGGGGCATGTT 143
    :  |||||||||||:::|::|||||||||
187 sTyrPheArgArgLeuAspGlyArgLeu 196
```

```
seq_name: SwissProt_39:FIMC_BORPE

seq_documentation_block:
ID   FIMC_BORPE     STANDARD;      PRT;    873 AA.
AC   P33410;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   OUTER MEMBRANE USHER PROTEIN FIMC PRECURSOR.
GN   FIMC OR FHAA.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WELLCOME 28;
RX   MEDLINE=93078620; PubMed=1360139;
RA   Willems R.J.L., van der Heide H.G.J., Mooi F.R.;
RT   "Characterization of a Bordetella pertussis fimbrial gene cluster
RT   which is located directly downstream of the filamentous
RT   haemagglutinin gene.";
RL   Mol. Microbiol. 6:2661-2671(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOHAMA I;
RX   MEDLINE=92371423; PubMed=1354611;
RA   Locht C., Geoffroy M.C., Renauld G.;
RT   "Common accessory genes for the Bordetella pertussis filamentous
RT   hemagglutinin and fimbriae share sequence similarities with the papC
RT   and papD gene families.";
RL   EMBO J. 11:3175-3183(1992).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions. There are no restrictions on its
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
DR   EMBL; X64876; CAA46090.1; -.
DR   EMBL; X66729; CAA47266.1; -.
DR   PIR; S21574; S21574.
DR   PIR; S25193; S25193.
DR   InterPro; IPR000015; -.
DR   Pfam; PF00577; Usher; 1.
DR   PROSITE; PS01151; FIMBRIAL_USHER; 1.
KW   Outer membrane; Transmembrane; Fimbria; Transport; Signal.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16    873       OUTER MEMBRANE USHER PROTEIN FIMC.
FT   DISULFID    815    838       POTENTIAL.
FT   CONFLICT    744    744       G -> A (IN REF. 2).
SQ   SEQUENCE   873 AA;  93369 MW;  ACB8D807838BCB22 CRC64;

alignment_scores:
            Quality:      69.00                  Length:      123
            Ratio:        0.986                  Gaps:          4
Percent Similarity:       56.911     Percent Identity:     25.203

alignment_block:
US-09-540-234-1/rev x FIMC_BORPE  ..

Align seg 1/1   to: FIMC_BORPE    from: 1   to: 873

 408 CCACTACAAAAGGGAAACGAAGTACTGTAGTACACAAGGTGTCCCCAGACCA   359
     |||:|||::||::|||||:||||  ::  |||||   :|  :|  :|| ||||
 597 ProValGlnGlnSerValAsnTyrSerGlyMetSerGlyLymGluArgAspGl   613

 358 GGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCCAAGGAAGTCTTCGAG   309
     |     |  :: |||    ||||||:||::|||:::  ::||
 613 nTyrThrTyrGlyLeuGlyValGlnArgAlaGlyThrSerAlaGlnTyrA   630

 308 TGCTGCACGGTAAACACAGCACTCAAATGGCTCCAAGCGGAGTACAGCAAGG   259
     :||:||:|||:   ||:::|||   :|:::|  ||| |||   |||| ||
 630 laLeuAsnGlySerTrpSerGlyThrTyrGlyGluValSer.GlyGlnLe   646

 258 AAGTCATCAAGGCCTACACAGTAAACAGCACACGAATGGCTCCGATAACA   209
     ::|||     ::|||:::  ::  |||||||||  ||  |||||  ||
 646 uThrHisGlyArgSerTyrSerGlnTyrGlnIleAsnGlySerGlyGlyL   663

 208 TCGACGCC......................ATTGGCGTT   192
     ::                                ||:||:|
 663 euValAlaHisAlaGlyGlyValThrPheGlyGlnTyrGlnAlaGlyTyrThr   679

 191 GTTAGTCTTGTACAAACTACGGCTTCACCG............ACGTC   157
     :::|||:||:|: |||::                         ::|||
 680 IleGlyLeuIleGlnAlaGlnAlaAlaGlyAlaAlaLysValValAsnTh   696

 156 TCGAGGGGCATGTTTGAGCAGCGACCCTCAAAAGGGCAGGGACTGTTTTCCA   107
     :||::|||||  :||  |||      :||||   ::|||  |||||| ||||
 696 rArgAsnAlaAlaValAspArg.....SerGlyTyrGlyLeuValSerL   711

 106 TTATGCCGTGCCAGCTA   90
     ::    |  ::|||
 711 euThrProTyrSerLeu   716
```

seq_name: SwissProt_39:M84D_DROME

seq_documentation_block:
```
ID   M84D_DROME    STANDARD;         PRT;         68 AA.
AC   Q01645;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   MALE SPECIFIC SPERM PROTEIN MST84DD.
GN   MST84DD.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=92102953; PubMed=1684716;
RA   Kuhn R., Kuhn C., Boersch D., Glaetzer K.H., Schaefer U.,
RA   Schaefer M.;
RT   "A cluster of four genes selectively expressed in the male germ line
RT   of Drosophila melanogaster.";
RL   Mech. Dev. 35:143-151(1991).
CC   -!- TISSUE SPECIFICITY: TESTIS.
CC   -!- DEVELOPMENTAL STAGE: PRIMARY SPERMATOCYTES.
CC   -!- DOMAIN: THIS PROTEIN IS MOSTLY COMPOSED OF REPETITIVE C-G-P
CC       MOTIFS.
CC   -!- SIMILARITY: BELONGS TO THE MST(3)CGP FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X67703; CAA47940.1; -.
DR   HSSP; P01180; 1NPO.
DR   FlyBase; FBgn0004175; Mst84Dd.
KW   Spermatogenesis; Repeat; Multigene family.
SQ   SEQUENCE   68 AA;  6480 MW;  2F2BD26128DE3DEF CRC64;

alignment_scores:
            Quality:      68.50                  Length:       33
            Ratio:        3.806                  Gaps:          3
Percent Similarity:       54.545     Percent Identity:     48.485

alignment_block:
US-09-540-234-1 x M84D_DROME   ..

Align seg 1/1   to: M84D_DROME    from: 1   to: 68

 229 TGCTGTTTACTGTGTAGCGCTTGATGACTTCCTTGCGTGTACTCCCGCTTGG   278
     |||||| |||   |||   ::| ||| :|  |||||||||  ||
  20 CysCysGlyProCysGly...........ProCysCysGlyProCysGl   33

 279 AGCCATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGC   325
     |||||||  :::|||  | |||||  |||  :|  ||||||   ||||||
  33 y.Pro.....CysCysGlyProCysGlyProArgCysGlyProCys   46
```

seq_name: SwissProt_39:TBA1_PELFA

seq_documentation_block:
```
ID   TBA1_PELFA    STANDARD;         PRT;        453 AA.
AC   Q40831;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TUBULIN ALPHA-1 CHAIN.
GN   TUBA1.
OS   Pelvetia fastigiata.
OC   Eukaryota; stramenopiles; Phaeophyceae; Fucales; Fucaceae; Pelvetia.
OX   NCBI_TaxID=48072;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Coffman H.R., Kropf D.L.;
RT   "The brown alga, Pelvetia fastigiata, expresses two alpha-tubulin
RT   sequences.";
RL   (In) Plant Gene Register PGR97-019.
CC   -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC       BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC       CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC   -!- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC   -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U58641; AAB68031.1; -.
DR   InterPro; IPR000217; -.
DR   InterPro; IPR002252; -.
DR   InterPro; IPR003008; -.
DR   Pfam; PF00091; tubulin; 1.
DR   PRINTS; PR01161; TUBULIN.
DR   PRINTS; PR01162; ALPHATUBULIN.
DR   PROSITE; PS00227; TUBULIN; 1.
KW   Microtubules; GTP-binding; Multigene family.
FT   NP_BIND   142   148   GTP (POTENTIAL).
FT   SITE   453   453   INVOLVED IN POLYMERIZATION.
SQ   SEQUENCE   453 AA; 49937 MW; 65DF46C1E2981ED4 CRC64;

alignment_scores:
         Quality: 68.00   Length: 80
         Ratio: 1.889   Gaps: 3
Percent Similarity: 45.000   Percent Identity: 28.750

alignment_block:
US-09-540-234-1 x TBA1_PELFA

Align.seg 1/1 to: TBA1_PELFA   from: 1 to: 453

55 GGACAAAGACACCTCCAAGCCGGACACGGC.................TACTGTAG 92
      |||||| ::||::||||::::::: ||||||
10 GlyGlnAlaGlyIleGlnValGlyAsnAlaCysTrpLeuTyrCysLe 26

93 CTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTTGAG.......... 132
      ||||||||:: |||:::: ||| ::::   :::
26 uGluHisGlyIleGlnProAsnGlyGlnMetProAlaAsnAsnThrThrG 43

133 ....GGTCGTCTC 141
                                                    |||
43 lyGlyGlyAspAspAlaPheAsnThrPhePheSerGluThrGlyAlaGly 59

142 AAACATGCCCCTCGAGACGTCGGT..........GAAGCCGTAGTTTGTAC 182
      ||||||::||||| ||| || :::|||||
60 LysHisValProArgAlaValTyrValAspLeuGluProThrValCysAs 76

183 AAGACTAACAACGCCAATGGCGTCGATGTTATCGGAGCCA 222
      :::::: ||| ::||| ::|||
76 pGluValArgThrGlySerTyrArgGlnLeuTyrHisPro 89

seq_name: SwissProt_39:YK25_CAEEL

seq_documentation_block:
ID   YK25_CAEEL   STANDARD;   PRT;   789 AA.
AC   P34332;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   HYPOTHETICAL 87.8 KDA PROTEIN C14B9.5 IN CHROMOSOME III.
GN   C14B9.5.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith A., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L15188; AAA27949.1; -.
DR   PIR; S44759; S44759.
DR   WormPep; C14B9.5; CE00076.
KW   Hypothetical protein.
SQ   SEQUENCE   789 AA; 87776 MW; 683S5D70D8F79DC CRC64;

alignment_scores:
         Quality: 68.00   Length: 123
         Ratio: 1.015   Gaps: 6
Percent Similarity: 54.472   Percent Identity: 26.829

alignment_block:
US-09-540-234-1/rev x YKZ5_CAEEL

Align seg 1/1 to: YKZ5_CAEEL from: 1 to: 789

399 AGGGAAACGAAGTACTGTAGTACAACAGGTGTCCCCAGACCAGGCAGCAAA 350
      |||::::::|||| :: |||::::||||| ||||||||
188 ArgGlnAlaGlnTyrProAlaLeuGlnAlaSerProHisGlnAlaAla.. 203

349 CAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTGCTGCACG 300
      ::::: |||||| ||||| ::|||::: :||
204 ...IleIleGlnHisGlnGlnGlnGlnAlaPheLysLysLeuGln. 218

299 GTAAACGACACTCAAATGGCTCCAGCGAGTACAGCAAGGAAGTCATCA 250
      :::||||||::||||||:::::::||||||
219 ..GlnGlnHisGlnAspAlaGlnLysArgLysLeuGluAlaProValSer 234

249 AGCGGCTACACAGTAA...................... 234
      |||||||| :::
235 SerAlaThr.ProGlnValLysArgIleAlaLeuSerGlnSerSerProV 251

233 ............CAGCACACGAATGGCTCC......GATAACATCGAC 204
                   |||:::: |||:::
251 alGlnArgPheSerGlnHisProAsnGlyThrLeuProHisAsnLeu.. 266

203 GCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTCTCG 154
      |||::   ::   :::|| ::||| ::||
267 AlaValValTyrGlnAsnLysMetValCysValArgThrProGlnAspAr 283

153 AGGGGCATGTTTGAGACGACCCTCAAAAGGGCAGGGACTGTTTTCCATTA 104
      |||:::::   :::||| ::||| :::
283 gGluLysLeuLeuGlnGluGlnGlnGlnLeuAsnGlnTyrLeuPheAlaAlaG 300

103 TGCCGTGCCAGCTACAG 87
      :::||||||
300 lnGlnGlnHisLeuGln 305

seq_name: SwissProt_39:COX1_ACACA

seq_documentation_block:
ID   COX1_ACACA   STANDARD;   PRT;   873 AA.
AC   Q37370;

```
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I+II (EC 1.9.3.1).
GN   COX1/2.
OS   Acanthamoeba castellanii (Amoeba).
OG   Mitochondrion.
OC   Eukaryota; Acanthamoebidae; Acanthamoeba.
OX   NCBI_TaxID=5755;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 30010 / NEFF;
RX   MEDLINE=95147275; PubMed=7844823;
RA   Burger G., Plante I., Lonergan K.M., Gray M.W.;
RT   "The mitochondrial DNA of the amoeboid protozoon, Acanthamoeba
RT   castellanii: complete sequence, gene content and genome
RT   organization.";
RL   J. Mol. Biol. 245:522-537(1995).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HEME-COPPER
CC       RESPIRATORY OXIDASE FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CYTOCHROME C
CC       OXIDASE SUBUNIT 2 FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U12386; AAD11820.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR001505; -.
DR   InterPro; IPR002428; -.
DR   InterPro; IPR002429; -.
DR   Pfam; PF00115; COX1; 1.
DR   Pfam; PF00116; COX2; 3.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PRINTS; PR01166; CYCOXIDASEII.
DR   PROSITE; PS00077; COX1; 1.
DR   PROSITE; PS00078; COX2; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   DOMAIN        1    474       COX1.
FT   DOMAIN      475    843       COX2.
FT   TRANSMEM     29     49       POTENTIAL.
FT   TRANSMEM     70     90       POTENTIAL.
FT   TRANSMEM     91    111       POTENTIAL.
FT   TRANSMEM    114    134       POTENTIAL.
FT   TRANSMEM    159    179       POTENTIAL.
FT   TRANSMEM    197    217       POTENTIAL.
FT   TRANSMEM    248    268       POTENTIAL.
FT   TRANSMEM    280    300       POTENTIAL.
FT   TRANSMEM    323    343       POTENTIAL.
FT   TRANSMEM    351    371       POTENTIAL.
FT   TRANSMEM    390    410       POTENTIAL.
FT   TRANSMEM    427    447       POTENTIAL.
FT   TRANSMEM    469    489       POTENTIAL.
FT   TRANSMEM    536    556       POTENTIAL.
FT   TRANSMEM    585    605       POTENTIAL.
FT   TRANSMEM    656    676       POTENTIAL.
FT   TRANSMEM    845    865       POTENTIAL.
FT   METAL        75     75       IRON (HEME A) (PROBABLE).
FT   METAL       254    254       COPPER B (PROBABLE).
FT   METAL       258    258       COPPER B (PROBABLE).
FT   METAL       303    303       COPPER B (PROBABLE).
FT   METAL       304    304       COPPER B (PROBABLE).
FT   METAL       389    389       IRON (HEME A3) (PROBABLE).
FT   METAL       391    391       IRON (HEME A) (PROBABLE).
FT   METAL       801    801       COPPER A (PROBABLE).
FT   METAL       836    836       COPPER A (PROBABLE).
FT   METAL       840    840       COPPER A (PROBABLE).
FT   METAL       844    844       COPPER A (PROBABLE).
SQ   SEQUENCE   873 AA;  99214 MW;  0BC820F0162B8EB3 CRC64;
```

```
alignment_scores:
                   Quality:   68.00            Length:  133
                     Ratio:    0.986              Gaps:    7
       Percent Similarity:   51.880   Percent Identity:  27.068

alignment_block:
US-09-540-234-1 x COX1_ACACA

Align seg 1/1  to: COX1_ACACA      from:  1  to: 873

  92 GCTGGCACGGCATAATGAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTC 141
     ||||||||||||  |||  |||||  |   ||| |||     :::|||||
 136 AlaGlyThrGly...TrpThrValTyrProLeuSerIleValAl 151

 142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
     ||||                                     :::   :::
 151 aHisSer.....GlyGlySerVal.AspLeuAlaIlePheSerLeuHis 165

 192 AACGCCAATGGCGTCGATGTATCGGAGCCATTCGTGTCGTGTTTACTGT 241
     ||||||  |||   :::||||||||||||  |||||  :::  ||||::|
 166 LeuAlaGlyIleSerSerLeuLeuGlyAlaIleAsnPheIleThrThrI 182

 242 GTAGCGCTTG..................ATGACTTCCTTGCTGCTACTCCG 273
     : :::|||||                  || |||||  ||||  |||||||
 182 ePheAsnMetArgValProGlyLeuSerMetHisLysLeuProLeuPheV 199

 274 CTTGGAGCCATTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTT 323
     ||||||| |||  |||||||::::|||||||||||  :::::   |||
 199 aIrpSerValLeuIleThrAlaPheLeuLeuLeuPheSer......Leu 213

 324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG................ 356
     ||||||||||::::    |||||||  |||||||
 214 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 230

 357 ..............................CCTGCTCGGGACACCTGTTACACAGTA 387
                                   |||   |||:::||  |||::::
 230 nThrSerPhePheAspProSerGlyGlyGlyAspProIleLeuTyrGlnH 247

 388 CTTCGTTT....................CCCTTTGTAGTGGTACTACTTCTT 419
     |||      ||                 ||| |||  ::::||||:::|||
 247 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeu 262
```

```
seq_name: SwissProt_39:COX1_PHYME

seq_documentation_block:
ID   COX1_PHYME     STANDARD;      PRT;   492 AA.
AC   Q02211;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1.
OS   Phytophthora megasperma (Potato pink rot fungus).
OG   Mitochondrion.
OC   Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
OC   Phytophthora.
OX   NCBI_TaxID=4788;
```

```
183 AAGACTAACAACGCCAATGGCTCGAT................... 209
    || :::::::::::    ::!!!!!!
145 ln.AlaHisSerGlyProSerValAspLeuAlaIlePheSerLeuHisLe 161

210 ...........GTTATCGGAGCCATTCGTGTGCTGTTACTGTGT 243
                :||||||||| ::: :|| ||:::
161 uThrGlyIleSerSerLeuLeuGlyAlaIleAsnPheIleSerThrIleT 178

244 ACGGCTGATGACT...........TCCTTGCTGACTCCGCT 275
    ::::::::::::   :::                    |||
178 yrAsnMetArgAlaProGlyLeuSerPheHisArgLeuProLeuPheVal 194

276 TGGAGCCATTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
    |||||  |||    ::::::::     ||    ||:::::::!:||
195 TrpSerValLeuIleThrAlaPheLeuLeuLeuLeuThr......LeuPr 209

326 TGTACTCGCTGGCAGCATAGTCTGTTTGCTG.............. 356
    |||||||||||| :::::::    ||||||
209 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnLeuAsnT 226

357 .................CCTGGTCTGGGGACCACCTGTGTACTACAGTACT 389
                       ||| |||||||   :::||   |||:::
226 hrSerPheTyrAspProSerGlyGlyAspProValLeuTyrGlnHis 242

390 TCGTT.........CCCTTGTAGTGGTACTACTTCTT 419
    ||| |||      ||:::::!:|||
243 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeu 257
```

seq_name: SwissProt_39:COX1_LOCMI

seq_documentation_block:
```
ID   COX1_LOCMI     STANDARD;      PRT;    512 AA.
AC   Q36421;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Locusta migratoria (Migratory locust).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Orthopteroidea; Orthoptera; Caelifera;
OC   Acridomorpha; Acridoidea; Acrididae; Locusta.
OX   NCBI_TaxID=7004;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96139026; PubMed=8587138;
RA   Flook P.K., Rowell C.H.F., Gellissen G.;
RT   "The sequence, organization, and evolution of the Locusta migratoria
RT   mitochondrial genome.";
RL   J. Mol. Evol. 41:928-941(1995).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=69ST;
RX   MEDLINE=90291515; PubMed=2357738;
RA   Shumard-Hudspeth D.S., Hudspeth M.E.;
RT   "Genic rearrangements in Phytophthora mitochondrial DNA.";
RL   Curr. Genet. 17:413-415(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=69ST;
RA   Sachay D.J., Hudspeth D.S., Nadler S.A., Hudspeth M.E.;
RT   "Oomycete mtDNA: Phytophthora genes for cytochrome c oxidase use an
RT   unmodified genetic code and encode proteins most similar to plants.";
RL   Exp. Mycol. 17:7-23(1993).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
```

```
DR   EMBL; L04457; AAA32023.2; -.
DR   HSSP; P00396; 1OCC.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        70     70       IRON (HEME A) (PROBABLE).
FT   METAL       249    249       COPPER B (PROBABLE).
FT   METAL       253    253       COPPER B (PROBABLE).
FT   METAL       298    298       COPPER B (PROBABLE).
FT   METAL       299    299       COPPER B (PROBABLE).
FT   METAL       384    384       IRON (HEME A3) (PROBABLE).
FT   METAL       386    386       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   492 AA;  54325 MW;  9214D85A2EE32021 CRC64;
```

alignment_scores:
```
          Quality:   67.00          Length:    149
            Ratio:    0.971            Gaps:      8
Percent Similarity:   46.309   Percent Identity:  27.517
```

alignment_block:
```
US-09-540-234-1 x COX1_PHYME ..

Align seg 1/1  to: COX1_PHYME  from: 1  to: 492

107 TGGAAAAGTCCCTGCCCTTTGAGGDYCGTCAAACATGCCCCTCGA 156
    |||   |||||||||| :::||||||||||:   :::||
112 TrpLeuLeuProAlaLeuLeuLeuLeuLeuValSerSerAlaIleValIleValValG1 128

157 GACG.........TCGGTGAAGCCCGTAGTTT.....GTAC 182
    |:::                  ::|||  ||| :::   ||||
128 uSerGlyAlaGlyThrGlyTrpThrValTyrProLeuSerSerValG 145
```

```
DR   EMBL; X80245; CAA56527.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane.
KW   Respiratory chain; Inner membrane.
FT   METAL        60     60       IRON (HEME A) (PROBABLE).
FT   METAL       239    239       COPPER B (PROBABLE).
FT   METAL       243    243       COPPER B (PROBABLE).
FT   METAL       289    289       COPPER B (PROBABLE).
FT   METAL       290    290       COPPER B (PROBABLE).
FT   METAL       375    375       IRON (HEME A3) (PROBABLE).
FT   METAL       377    377       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   512 AA;  56814 MW;  98031A229B00511C CRC64;

alignment_scores:
        Quality:     67.50          Length:    159
        Ratio:        0.877          Gaps:        9
Percent Similarity:  48.428   Percent Identity:  27.044

alignment_block:
US-09-540-234-1 x COX1_LOCMI

Align seg 1/1  to: COX1_LOCMI  from: 1  to: 512  ..

66  CCTCCAAGCCGGACACGGCTACTGTA.............GC 93
    ||||||||| |||| |||||
105 ProProSerLeuThrLeuLeuLeuMetSerSerValValAspAsnGlyAl 121

94  TGGCACGGCATATGAAAACAGTCCCTGCCCTTTGAGGGTGCGTCTCAA 143
    |||||||::: |||  ::|||  |||  |||||||||:::::::::
121 aGlyThrGly...TrpThrValTyrProProLeuAlaSerValIleAlaH 137

144 ACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAA 193
       ::|||||||  ::  :::::::::::   ::
137 isSer......GlyAlaSerVal.AspLeuAlaIlePheSerLeuHisLe 151

194 CGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGTGT 243
    |||||||||         ||||| ||::: ||  ::||:::
151 uAlaGlyValSerSerIleLeuGlyAlaIleAsnPheIleThrThrAlaI 168

244 AGCGCCGTTGATGACTTCCTTGCTG..............TACTCCGCT 275
    ::::::    :::::  ::                    ::|||
168 leAsnMetArgSerAsnAsnMetThrLeuAspGlnThrProLeuPheVal 184

276 TGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
    ||||||            |||        ||| :::
185 TrpSer......ValAlaIleThrAlaLeuLeuLeuLeuLeuSerLeuPr 199

326 TGTACTCGCTGGCAGCATAGTCTGTTTGCTG................ 356
    ||||:::  |||||||||:::||:::::||
199 oValLeuAlaGlyAlaIleThrMetLeuLeuThrArgAsnLeuAsnT 216

357 ...........CCTGGTCTCGGGACCACCTGTGTACTACAGTACT 389
               ||| |||::: :::||
216 hrSerPhePheAspProAlaGlyGlyAspProIleLeuTyrGlnHis 232

390 TCGTTT..............CCCTTTGTAGTGGTACTACTTCTT...... 419
    |||                 ||| ::||||||||:::|||
233 LeuPhePhePheGlyHisProGluValTyrIleLeuIleLeuProGl 249

420 ...........AGCCATGTACTC 431
               |||:|::|||:::||
249 yPheGlyIleIleSerHisIleVal 257

seq_name:  SwissProt_39:COX1_RHISA

seq_documentation_block:
ID   COX1_RHISA     STANDARD;      PRT;    512 AA.
AC   O99818;
```

```
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Rhipicephalus sanguineus (Brown dog tick).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC   Parasitiformes; Ixodida; Ixodidae; Rhipicephalus.
OX   NCBI_TaxID=34632;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99903443; PubMed=9866211;
RA   Black W.C. IV, Roehrdanz R.L.;
RT   "Mitochondrial gene order is not conserved in arthropods: prostrate
RT   and metastriate tick mitochondrial genomes.";
RL   Mol. Biol. Evol. 15:1772-1785(1998).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF081829; AAD05518.1; -.
DR   HSSP; P98002; 1AR1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002106; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        60     60       IRON (HEME A) (PROBABLE).
FT   METAL       239    239       COPPER B (PROBABLE).
FT   METAL       243    243       COPPER B (PROBABLE).
FT   METAL       289    289       COPPER B (PROBABLE).
FT   METAL       290    290       COPPER B (PROBABLE).
FT   METAL       375    375       IRON (HEME A3) (PROBABLE).
FT   METAL       377    377       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   512 AA;  57125 MW;  53D6374C4DF11AB6 CRC64;

alignment_scores:
        Quality:     67.00          Length:    133
        Ratio:        1.000          Gaps:        7
Percent Similarity:  50.376   Percent Identity:  27.068

alignment_block:
US-09-540-234-1 x COX1_RHISA

Align seg 1/1  to: COX1_RHISA  from: 1  to: 512  ..

92  GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    ||||||||||: ||  :::  ||| ||| |||  :::||
121 AlaGlyThrGly...TrpThrValTyrProProLeuSerSerLeuSe 136
```

```
142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    |:|::         ||||||  ::  ::|||::||:::  ::
136 rHis.....TyrGlyProSerVal.AspLeuAlaIlePheSerLeuHis 150

192 AACGGCCAATGGCGTCGATGTATCGGAGCCATTCGTGTGCTGTTTACTGT 241
    ||:::::    :||||:::::||||||||||||  :::  :|||::
151 LeuAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrThrIl 167

242 GTAGGCCTTG.............ATGACTTCCTTGCTGACTCCG 273
    ::::|||                |||
167 EValAsnMetArgSerIleGlyMetThrMetGluArgMetProLeuPheV 184

274 CTTGGAGCCATTGAGTGCTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTT 323
    ::||||| |||  :|||:            |||::||     |||
184 alTrpSerValLeuIleThrAlaIleLeuLeuLeuLeuSer......Leu 198

324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG................. 356
    ||||||||||||:|||:::|:  |||||||
199 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgArgAsnPheAs 215

357 ......CCTGTCTGGGGACACCTGTGTACTACAGTA 387
          |||  |||   |||:||||   ||::|||
215 nThrSerPheAspProSerGlyGlyGlyAspProIleLeuTyrArgGlnH 232

388 CTTCGTTT...........CCCTTTGTAGTGGTACTACTTCTT 419
    |||                ||| |||:|||||:||:||
232 isLeuPheTrpPheGlyHisProGluValTyrIleLeuIleLeu 247
```

seq_name: SwissProt_39:CIK3_MOUSE

CC   -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM
CC   ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED
CC   CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE
CC   MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH
CC   WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL
CC   GRADIENT.
CC   -!- SUBUNIT: HETEROTETRAMER OF POTASSIUM CHANNEL PROTEINS (PROBABLE).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC   -!- DOMAIN: THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE
CC   RATE OF INACTIVATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE
CC   IN MODULATION OF CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL
CC   TO SPECIFIC SUBCELLULAR COMPARTMENTS.
CC   -!- DOMAIN: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS
CC   CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
CC   EVERY THIRD POSITION.
CC   -!- SIMILARITY: THIS CHANNEL PROTEIN BELONGS TO THE DELAYED RECTIFIER
CC   CLASS. BELONGS TO SHAKER POTASSIUM CHANNEL SUBFAMILY.
CC   --

```
DR   EMBL; M30441; AAA39716.1; -.
DR   HSSP; P03621; 2IFN.
DR   MGD; MGI:96660; Kcna3.
DR   InterPro; IPR000636; -.
DR   InterPro; IPR003091; -.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR00169; KCHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Glycoprotein; Multigene family; Phosphorylation.
FT   TRANSMEM      188    206       SEGMENT S1.
FT   TRANSMEM      248    269       SEGMENT S2.
FT   TRANSMEM      281    301       SEGMENT S3.
FT   TRANSMEM      316    334       SEGMENT S4.
FT   TRANSMEM      351    370       SEGMENT S5.
FT   TRANSMEM      412    434       SEGMENT S6.
FT   CARBOHYD       62     62       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      231    231       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      232    232       N-LINKED (GLCNAC...) (POTENTIAL).
FT   MOD_RES       473    473       PHOSPHORYLATION (BY CAPK)
FT                                  (POTENTIAL).
SQ   SEQUENCE   528 AA;   58526 MW;  CBD65F15l5F9A731 CRC64;
```

```
alignment_scores:
         Quality:   67.00               Length:    38
           Ratio:    2.233                Gaps:     2
Percent Similarity:  78.947     Percent Identity: 47.368

alignment_block:
US-09-540-234-1/rev x CIK3_MOUSE

Align seg 1/1  to: CIK3_MOUSE   from: 1  to: 528

368 CCCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGA 319
    ||:::    ||:::||||       |||||   |||||   ||||||
210 ProGluPheArgAspGluLysAspTyrProAlaSerPro...SerGlnAs 225

318 AGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCGGAG 269
    :|||||||||:::    |||||||||        |||||||||||  |||||||
225 pValPheGluAlaAlaAsn...AsnSerThrSerGlyAlaProSerGlyA 241

268 TACACGCAAGGAGT 255
    :::::::::|||
241 laSerSerPheSer 245
```

seq_name: SwissProt_39:MBHL_ALCHY

CC   -!- FUNCTION: THIS ENZYME RECYCLES THE H(2) PRODUCED BY NITROGENASE TO
CC   INCREASE THE PRODUCTION OF ATP AND TO PROTECT NITROGENASE AGAINST
```

CC   INHIBITION OR DAMAGE BY O(2) UNDER CARBON- OR PHOSPHATE-LIMITED
CC   CONDITIONS.
CC   -!- CATALYTIC ACTIVITY: 2 REDUCED FERREDOXIN + 2 H(+) = 2 OXIDIZED
CC   FERREDOXIN + H(2).
CC   -!- COFACTOR: CONTAINS NICKEL AND IRON-SULFUR.
CC   -!- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.
CC   -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND.
CC   -!- SIMILARITY: BELONGS TO THE [NIFE]/[NIFESE] HYDROGENASE LARGE
CC   SUBUNIT FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; S56898; AAB25780.1; ALT_SEQ.
DR   PIR; JH0776; JH0776.
DR   HSSP; P12944; 1FRV.
DR   InterPro; IPR001501; -.
DR   Pfam; PF00374; NiFeSe_Hases; 1.
DR   PROSITE; PS00507; NI_HGENASE_L_1; FALSE_NEG.
DR   PROSITE; PS00508; NI_HGENASE_L_2; 1.
KW   Oxidoreductase; Membrane; Nickel; Plasmid.
FT   METAL   75   75   NICKEL (POTENTIAL).
FT   METAL   78   78   NICKEL (POTENTIAL).
FT   METAL   600   600   NICKEL (POTENTIAL).
FT   METAL   603   603   NICKEL (POTENTIAL).
SQ   SEQUENCE 621 AA; 68924 MW; 906AC4414285D4A3 CRC64;

alignment_scores:
     Quality: 67.00   Length: 88
     Ratio: 1.634   Gaps: 6
Percent Similarity: 46.591   Percent Identity: 31.818

alignment_block:
US-09-540-234-1 x MBHL_ALCHY

Align seg 1/1   to: MBHL_ALCHY   from: 1 to: 621

 37 TGGGTCATCTGAAGGAAAGGAAACGACACCTCCAGCCGGACACGGCTA 86
    ||| :::::: ||||||| ||| :::|||
519 TrpProLysGluAlaLysGlyValGlyHisValAlaAlaProArgGlyAl 535

 87 CTGTAGC...TGG.........CACGGCATAATGGAAAAACAGTCCCT 121
    ::||::: ||| ||| :::||||
535 aCysGlyHisTrpIleArgIleLysAspGlyLysIleGluAsnTyrGlnC 552

122 GCCCTTTGAGGGTCGTCTCAAACATGCCCCTCGAGAC........... 159
    || ::|||||||||
552 ysValValProThrThrTrpAsnGlySerProArgAspSerLysGlyGln 568

160 GTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCGTCGAT 209
    :::||| :::::::::::::::
569 IleGly...AlaPheGluAlaSerLeuMetAsnThrProMetAlaLysPr 584

210 GTTATCGGAGCCA............................TTCGTCGTGCT 232
    ::||||| |||
584 o...GluGluProValGluIleLeuArgThrValHisSerPheAspProC 600

233 GTTTACTGTGTAGC 246
    ||||| |||||
600 ysLeuAlaCysSer 604

seq_name: SwissProt_39:FABB_HAEIN

seq_documentation_block:
ID   FABB_HAEIN   STANDARD;   PRT;   406 AA.
AC   P43710;
DT   01-NOV-1995 (Rel. 32, Created)

DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   3-OXOACYL-[ACYL-CARRIER-PROTEIN] SYNTHASE I (EC 2.3.1.41) (BETA-
DE   KETOACYL-ACP SYNTHASE I) (KAS I).
GN   FABB OR HI1533.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RD / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus
RT   influenzae Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: CATALYZES THE CONDENSATION REACTION OF FATTY ACID
CC   SYNTHESIS BY THE ADDITION TO AN ACYL ACCEPTOR OF TWO CARBONS
CC   FROM MALONYL-ACP. SPECIFIC FOR ELONGATION FROM C-10 TO UNSATURATED
CC   C-16 AND C-18 FATTY ACIDS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ACYL-[ACYL-CARRIER PROTEIN] + MALONYL-[ACYL-
CC   CARRIER PROTEIN] = 3-OXOACYL-[ACYL-CARRIER PROTEIN] + CO(2) +
CC   [ACYL-CARRIER PROTEIN].
CC   -!- PATHWAY: FATTY ACID BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC   -!- SIMILARITY: BELONGS TO THE BETA-KETOACYL-ACP SYNTHASES FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U32829; AAC23183.1; -.
DR   TIGR; HI1533; -.
DR   InterPro; IPR000794; -.
DR   Pfam; PF00109; ketoacyl-synt; 1.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW   Fatty acid biosynthesis; Transferase; Acyltransferase.
FT   ACT_SITE 162   162   BY SIMILARITY.
SQ   SEQUENCE 406 AA; 42649 MW; A17B8673156FDDAF CRC64;

alignment_scores:
     Quality: 66.50   Length: 118
     Ratio: 1.090   Gaps: 5
Percent Similarity: 51.695   Percent Identity: 25.424

alignment_block:
US-09-540-234-1/rev x FABB_HAEIN

Align seg 1/1   to: FABB_HAEIN   from: 1 to: 406

440 AACAGACACGAGTACATGCTAAGAAGTAGTACCACTACAAAGGGAAACG 391
    ||| ||||||| :::||| :::::::| |||
 55 AsnProSerGluHisIleAspArgLysValPheArgPheMetGly...... 69

390 AAGTACTGTAGTACACAGGTGTCCCAGACCAGGCAGCAAAACAGACTATG 341
                     ||| |||
 70 ....AspAlaA 72

```
340 CTGCCAGCCAGTAC.....AGCAAGAGGAAGTCTTCGAGTGCTGCACGGTA 297
        |||           ::|||||||::::|||    ::  ::
 72 laAlaTyrAlaTyrLeuSerMetArgGluAlaIleGluAspAlaGlyLeu 88

296 AACAGCACTCAAATGGCTCCAAGC...............GGAGTACA 265
        ::::: |||::|||   |||::::              |||:|:
 89 ThrGluAspGlnValSerAsnAspArgThrGlyLeuValIleGlyAlaGl 105

264 GCAAGGAAGT...CATCAAGCGCTACACAGTAAACAGCACGAATGGCT 218
        ||||||||      |||::  :::     :::::::|||
105 yThrGlySerAlaHisAsnGlnLeuValAlaCysAspAlaValArgGlyP 122

217 CCGATAACATCGACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCT 168
        ::::: |||||||| |||||||||     ::::::|||::|| |||
122 roArgGlyValLysAlaIleGlyProTyrAlaValThrLysThrMetAla 138

167 TCACCGACCTCGAGGGGCATGTTGAGACGACCCTCAAAAGGGCAGG 118
        ||| ||||||    ||||||||     ::::::|||
139 SerSerValSer......AlaCysLeuAlaThrProTyrLysIleArgGl 153

117 ACTG 114
    |:::
153 yVal 154
```

seq_name: SwissProt_39:FRA_DROME

```
seq_documentation_block:
ID   FRA_DROME     STANDARD;        PRT;      595 AA.
AC   P21525;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   TRANSCRIPTION FACTOR DFRA (FOS-RELATED ANTIGEN) (AP-1) (KAYAK
DE   PROTEIN).
GN   KAY OR FRA.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=90337318; PubMed=2116361;
RA   Perkins K.K., Admon A., Patel N., Tjian R.;
RT   "The Drosophila Fos-related AP-1 protein is a developmentally
RT   regulated transcription factor.";
RL   Genes Dev. 4:822-834(1990).
CC   -!- FUNCTION: DEVELOPMENTALLY REGULATED TRANSCRIPTION FACTOR AP-1.
CC       BINDS AND RECOGNIZES THE ENHANCER DNA SEQUENCE: TGA(C/G)TCA. MAY
CC       PLAY A ROLE IN THE FUNCTION OR DETERMINATION OF A PARTICULAR
CC       SUBSET OF CELLS IN THE DEVELOPING EMBRYO. IT IS ABLE TO CARRY OUT
CC       ITS FUNCTION EITHER INDEPENDENTLY OF OR IN CONJUNCTION WITH DJRA.
CC   -!- SUBUNIT: MAY FORM DIMER OF IDENTICAL CHAINS AND MAY ALSO INTERACTS
CC       WITH DJRA TO FORM AN HETERODIMER. DFRA-DJRA COMPLEX IS BOUND MORE
CC       STABLY TO THE AP-1 SITE THAN EITHER OF THE TWO PROTEINS ALONE.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR.
CC   -!- TISSUE SPECIFICITY: CELLS TYPE OF THE EMBRYO THAT ARE INVOLVED IN
CC       THE DEVELOPMENT OF THE HEAD AND NERVOUS SYSTEM.
CC   -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. FOS SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X54143; CAA38082.1; -.
DR   PIR; A35847; A35847.
DR   HSSP; P01100; 1FOS.
DR   TRANSFAC; T01997; -.
DR   FlyBase; FBgn0001297; kay.
```

```
DR   InterPro; IPR001871; -.
DR   Pfam; PF00170; bZIP; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT   DNA_BIND    260    279       BASIC MOTIF.
FT   DOMAIN      286    314       LEUCINE-ZIPPER.
SQ   SEQUENCE   595 AA;  62837 MW;  C4D81F99582D38DE CRC64;
```

```
alignment_scores:
            Quality:   66.50        Length:      69
            Ratio:      1.357        Gaps:        2
Percent Similarity: 71.014    Percent Identity: 26.087
```

```
alignment_block:
US-09-540-234-1/rev x FRA_DROME  ..

Align seg 1/1 to: FRA_DROME from: 1 to: 595

357 GCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAGCAAGGAAGTCTTCGAGT 308
    |||:|||||||  |||||||    ||||::::::::::::||   ||| |||||
319 AlaThrHisArgAlaThrCysGlnLysIleArgSerAspMetLeuSerVa 335

307 GCTGCACGGTAAACAGCAGCACTCAAATGGCTCCAAGCCGGAGTACAGCAAGGA 258
    |||:|| ||| |||::      ::||||||||::::|||||:::    ::::
335 lVal.ThrCysAsnGlyLeu...IleAlaProAlaGlyLeuLeuSerAla 350

257 AGTCATCAAGCGCTACACAGTAAACAGCACAGCACACGAATGGCTCCGATAACAT 208
    :::  :::::||  ::::::|||    ||| |||     |||::||| |||||
351 GlySerSerGlySerGlyAlaSerSerHisHisAsnHisAsnSerAsnAs 367

207 CGACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGT 158
    ||| |||:::      ||||||::||||||||     :||||||||
367 pSerSerAsnGlyThrIleThrGlyMetAspAlaThrLeuAsnSerThrG 384
```

```
157 CTCGA 153
    ::|||
384 lyArg 385
```

seq_name: SwissProt_39:PRIA_BACSU

```
seq_documentation_block:
ID   PRIA_BACSU    STANDARD;        PRT;      805 AA.
AC   P94461; O34941;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
GN   PRIA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Foulger D., Errington J.;
RT   "DNA sequence of a 28 Kbp segment of DNA from the spoVM region of
RT   Bacillus subtilis.";
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 230-895 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=97240812; PubMed=9086272;
RA   Mazel D., Coic E., Blanchard S., Saurin W., Marliere P.;
RT   "A survey of polypeptide deformylase function throughout the
RT   eubacterial lineage.";
RL   J. Mol. Biol. 266:939-949(1997).
CC   -!- FUNCTION: RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX SSDNA.
CC       THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS PRIB AND
CC       PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE SUBSEQUENT
CC       ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN FUNCTIONS AS A
CC       HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Y13937; CAA74261.1; -.
DR   EMBL; Z99112; CAB13444.1; -.
DR   EMBL; Y10304; CAA71348.1; -.
DR   Subtilist; BG11963; priA.
KW   DNA replication; DNA-binding; ATP-binding; Helicase; Primosome;
KW   Zinc-finger.
FT   NP_BIND     295    302      ATP (POTENTIAL).
FT   SITE        391    394      DEEH BOX.
FT   ZN_FING     510    522      C4-TYPE (POTENTIAL).
FT   ZN_FING     537    553      C4-TYPE (POTENTIAL).
FT   CONFLICT    232    232      A -> V (IN REF. 2).
FT   CONFLICT    691    698      FYQHEMAH -> VLSSMKWRT (IN REF. 2).
SQ   SEQUENCE   805 AA;  91353 MW;  7E0D2970805AC948 CRC64;
```

```
alignment_scores:
        Quality:  66.50          Length:      128
        Ratio:    1.167          Gaps:          6
Percent Similarity: 44.531   Percent Identity: 24.219

alignment_block:
US-09-540-234-1/rev x PRIA_BACSU

Align seg 1/1  to: PRIA_BACSU  from: 1  to: 805

434 CACGAGTACACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGAAGTAC 385
    |||::|||        |||::|||   |||  |||||||||:::||||
529 HisArgTyrGlyGlnArgLeuLysCysHisTyrCysGlyHisGluGluPr 545

384 TGTAGTACACAGGTGTCCCCAGCACAGGCAGCAAACAGACTATGCTGCCA 335
    |||  |||  ||||||:::                        :::||||
545 oValProHisThrCysProGlu...................CysAlaS 555

334 GCGAGTAC.................AGCAAGGAAGTCTTCGAGTGC 306
    |||||||:                  :::||||||||:|| |||
555 erGluHisIleArgPhePheGlyThrGlyThrGlnArgValGluGlu... 570

305 TGCACGGTAAACAGCACTCAAATGGCTCCAGCGGA................ 270
       ::: |||:||||::::     ||||||||
571 ....GluLeuThrLysValLeuProSerAlaArgValIleArgMe 584

269 .............GTACAGCAAGGAAGTCATCAAGCGCTACACAGT 237
                 ::::|||::|||:||  |||  |||
584 tAspValAspThrThrSerArgLysGlyAlaHisGluLysLeuLeuSerA 601

236 .............AAACAGCACACG 225
                ||||::||||:||||
601 laPheGlyGluGlyLysAlaAspIleLeuLeuGlyThrGlnMetIleAla 617

224 AATGGCTCCGAT........ACAATCGACGCGACCATTGGCCTTGGTTAGTCTTGT 181
    ::::|||    |||        :::|||||||||::||||
618 LysGlyLeuAspPheProAsnValThrLeuValGlyValLeuSerAlaAs 634

180 ACAAACTACGGCTTCACCGACCCTCCAGGGGCA 147
    |||  |||   |||  |||::|||
634 pThrThrLeuHisIleProAspPheArgSerAla 645
```

```
seq_name: SwissProt_39:AGSR_HUMAN

seq_documentation_block:
ID   AGSR_HUMAN     STANDARD;      PRT;    132 AA.
AC   O00253; Q15459;
DT   01-NOV-1997 (Rel. 35, Created)
```

```
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   AGOUTI RELATED PROTEIN PRECURSOR.
GN   AGRP OR ART OR AGRT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97230362; PubMed=9119224;
RA   Shutter J.R., Graham M., Kinsey A.C., Scully S., Luethy R.,
RA   Stark K.L.;
RT   "Hypothalamic expression of ART, a novel gene related to agouti, is
RT   up-regulated in obese and diabetic mutant mice.";
RL   Genes Dev. 11:593-602(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=97458244; PubMed=9311920;
RA   Ollmann M.M., Wilson B.D., Yang Y.K., Kerns J.A., Chen Y., Gantz I.,
RA   Barsh G.S.;
RT   "Antagonism of central melanocortin receptors in vitro and in vivo by
RT   agouti-related protein.";
RL   Science 278:135-138(1997).
RN   [3]
RP   DISULFIDE BONDS.
RX   MEDLINE=98393470; PubMed=9724530;
RA   Bures E.J., Hui J.O., Young Y., Chow D.T., Katta V., Rohde M.F.,
RA   Zeni L., Rosenfeld R.D., Stark K.L., Haniu M.;
RT   "Determination of disulfide structure in agouti-related protein (AGRP)
RT   by stepwise reduction and alkylation.";
RL   Biochemistry 37:12172-12177(1998).
RN   [4]
RP   STRUCTURE BY NMR OF 87-132.
RX   MEDLINE=99297561; PubMed=10371151;
RA   Bolin K.A., Anderson D.J., Trulson J.A., Thompson D.A., Wilken J.,
RA   Kent S.B.H., Gantz I., Millhauser G.L.;
RT   "NMR structure of a minimized human agouti related protein prepared by
RT   total chemical synthesis.";
RL   FEBS Lett. 451:125-131(1999).
CC   -!- FUNCTION: PLAYS A ROLE IN WEIGHT HOMEOSTASIS. MAY PLAY A ROLE IN
CC       THE REGULATION OF MELANOCORTIN RECEPTORS WITHIN THE HYPOTHALAMUS
CC       AND ADRENAL GLAND, AND THEREFORE IN THE CENTRAL CONTROL OF
CC       FEEDING.
CC   -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED PRIMARILY IN THE ADRENAL GLAND,
CC       SUBTHALAMIC NUCLEUS, AND HYPOTHALAMUS, WITH A LOWER LEVEL OF
CC       EXPRESSION OCCURRING IN TESTIS, LUNG, AND KIDNEY.
CC   -!- SIMILARITY: BELONGS TO THE AGOUTI FAMILY.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U88063; AAB52240.1; -.
DR   EMBL; U89485; AAB68621.1; -.
DR   MIM; 602311; -.
KW   Signal.
FT   SIGNAL       1     20        POTENTIAL.
FT   CHAIN       21    132        AGOUTI RELATED PROTEIN.
FT   DOMAIN      87    129        CYS-RICH.
FT   DISULFID    87    102
FT   DISULFID    94    108
FT   DISULFID   101    119
FT   DISULFID   105    129
FT   DISULFID   110    117
FT   CONFLICT     6      6        V -> L (IN REF. 2).
SQ   SEQUENCE   132 AA;  14440 MW;  1CCBE112C3EB10F5 CRC64;
```

```
alignment_scores:
    Quality:     66.00          Length:        84
    Ratio:        2.062          Gaps:           4
Percent Similarity:  38.095   Percent Identity: 23.810

alignment_block:
US-09-540-234-1 x AGSR_HUMAN  .

Align seg 1/1  to: AGSR_HUMAN  from: 1  to: 132

194 CGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGCTGTTTACTGTGT 243
    |||::::  ||||||||| |||::|:               |||::
 82 ArgSerSerArgArgCysValArgLeuHisGluSer........CysLe 95

244 AGCGCTTGATGACTTCCTTGCGTACTCCGCGTTGGAGCCATTTGAGTGCT 293
    ::::                                :::      ||||
 95 uGly..............................GlnGlnValProCysC 102

294 GTTTACCGTGCAGCACTCGAAGACTTCCTTGCTGCTGTACTCGCTGCAGCAT 343
    ||   ||||||:::
102 ysAspProCysAla......... 106

344 AGTCTGTTTGCTGCCTGGTCTGGGACACCTGTGTACTACAGTACTTCGT 393
             ||||||                :::::|||
107 ......................ThrCysTyrCysArgPhePheAs 114

394 TTCCCTTTGT............AGTGGTACTACTTCCTTAGCCATGTACTC 431
    ::::|||::                           |||||||:::|
114 nAlaPheCysTyrCysArgLysLeuGlyThrAlaMetAsnProCysSerA 131

432 GT 433
    ||
131 rg 131


seq_name: SwissProt_39:GBA5_DICDI

seq_documentation_block:
ID   GBA5_DICDI     STANDARD;      PRT;    347 AA.
AC   P34043;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   GUANINE NUCLEOTIDE-BINDING PROTEIN ALPHA-5 SUBUNIT (G ALPHA 5).
GN   GPAE.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96205045; PubMed=8620848;
RA   Hadwiger J.A., Natarajan K., Firtel R.A.;
RT   "Mutations in the Dictyostelium heterotrimeric G protein alpha
RT   subunit G alpha5 alter the kinetics of tip morphogenesis.";
RL   Development 122:1215-1224(1996).
RN   [2]
RP   SEQUENCE OF 205-322 FROM N.A.
RX   MEDLINE=91376114; PubMed=1910174;
RA   Hadwiger J.A., Wilkie T.M., Strathmann M., Firtel R.A.;
RT   "Identification of Dictyostelium G alpha genes expressed during
RT   multicellular development.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:8213-8217(1991).
CC   -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC       INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC       SIGNALING SYSTEMS.
CC   -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC       THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC   -!- MISCELLANEOUS: THERE ARE 8 KNOWN ALPHA SUBUNITS IN D.DISCOIDEUM.
CC   -!- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY. SUBFAMILY 3 (G(Q)).
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; U20806; AAB04097.1; -.
DR   PIR; B40990; B40990.
DR   HSSP; P10824; 1GDD.
DR   DictyDb; DD02026; gpaE.
DR   InterPro; IPR001019; -.
DR   Pfam; PF00503; G-alpha; 1.
DR   PRINTS; PR00318; GPROTEINA.
KW   GTP-binding; Transducer; Multigene family.
FT   NP_BIND       35     42       GTP (BY SIMILARITY).
FT   NP_BIND      195    199       GTP (BY SIMILARITY).
FT   NP_BIND      264    267       GTP (BY SIMILARITY).
FT   CONFLICT     248    263       IVNSHWFRNTAFIIFF -> LLIVIGLEIQHSLYFL (IN
FT                                 REF. 2).
FT   CONFLICT     274    274       A -> R (IN REF. 2).
FT   CONFLICT     284    286       AYT -> VYP (IN REF. 2).
FT   CONFLICT     299    299       K -> N (IN REF. 2).
FT   CONFLICT     309    309       N -> S (IN REF. 2).
FT   CONFLICT     320    320       I -> T (IN REF. 2).
SQ   SEQUENCE   347 AA;  40407 MW;  26A9A347BE103993 CRC64;


alignment_scores:
    Quality:     66.00          Length:        74
    Ratio:        1.650          Gaps:           2
Percent Similarity:  54.054   Percent Identity: 22.973

alignment_block:
US-09-540-234-1 x GBA5_DICDI  .

Align seg 1/1  to: GBA5_DICDI  from: 1  to: 347

191 CAACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGCTGTTTACTG 240
    :||||:|||||||  :|||::   ||:|| |||    :::::|||:|||:::
203 ArgArgLysTrpIleHisCysAspSerValThrAlaValIlePheCy 219

241 TGTAGCGCTTGATGACTTCCTTGCGTACTCCGCGTTGGAGCCATTTGAGT 290
    ||||||||||:::::::||:::::::    :::|||||||| |||
219 sValAlaLeuSerGluTyrAspGlnThrLeuArgGluGluSerGln. 235

291 GCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGCTGCTGTACTCGCTGCAG 340
                                       :::|||   :::||
236 ...................................AsnArgMetLys 239

341 CATAGTCTGTTT..........................GCTGCCTGGTCTGGGGA 369
    ::|||||||||                          :::   |||     ::
240 GluSerLeuMetLeuPheAspGluIleValAsnSerHisTrpPheArgAs 256

370 CACCTGTGTACTACAGTACTTC 391
    :|||:||| ::: :::|||
256 nThrAlaPheIleIlePhePhe 263


seq_name: SwissProt_39:COX1_PELSU

seq_documentation_block:
ID   COX1_PELSU     STANDARD;      PRT;    514 AA.
AC   O79672;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COI.
OS   Pelomedusa subrufa (African side-necked turtle).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Testudines; Pleurodira; Pelomedusidae; Pelomedusa.
OX   NCBI_TaxID=44522;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RA   Zardoya R.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AF039066; AAD05052.1; -.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        61     61       IRON (HEME A) (PROBABLE).
FT   METAL       240    240       COPPER B (PROBABLE).
FT   METAL       244    244       COPPER B (PROBABLE).
FT   METAL       290    290       COPPER B (PROBABLE).
FT   METAL       291    291       COPPER B (PROBABLE).
FT   METAL       376    376       IRON (HEME A3) (PROBABLE).
FT   METAL       378    378       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   514 AA;  56977 MW;  195105D81C814EB1 CRC64;

alignment_scores:
         Quality:              66.00          Length:         144
         Ratio:                 0.892         Gaps:             9
Percent Similarity:    51.389          Percent Identity:   27.083

alignment_block:
US-09-540-234-1 x COX1_PELSU

Align seg 1/1  to: COX1_PELSU   from: 1  to: 514

  92 GCTGGCACGGCATAATGGAAAACAGTCCTGCCCTTTGAGGGTCGTCTC 141
     ||||||||||::  |||  ::: |||  ||        :::||
 122 AlaGlyThrGly...TrpThrValTyrProProLeu.........AlaSe 134

 142 AAACATGCCC...CTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACT 188
     :||| |||::: ::||||||| :::|||::||:::|||::
 134 rAsnLeuAlaHisAlaGlyAlaSerVal.AspLeuAlaIlePheSerLeu 150

 189 AACAACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTCGTGTTTAC 238
     :::  |||::|||  |  || ||    :||  || |||||
 151 HisLeuAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrTh 167

 239 TCGTTAGGCCTTGATGACT.................TCCTTGCTGTACT 270
     ||||  :::::: ||| |                  :::||||
 167 rValIleAsnMetLysThrProAsnMetSerPheLeuAspMetProLeuP 184

 271 CCGCTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTC 320
     |||||||  :::::: ||| |               |||::
 184 heValTrpSerValLeuIleThrAlaIleLeuLeuLeuLeuSer...... 198
```

```
321 CTTGCTGCTACTCGCTGGCAGCAATAGTCTGTTGCTG............. 356
    |||  |||||||||::::|||::::|||  ||||||
199 LeuProValLeuAlaAlaGlyLyIleThrMetLeuLeuThrAspArgAsnLe 215

357 .............................CCTGGCTCGGGACACCTGTACTACA 384
    |||  :::                     |||  |||::: |||  |||
215 uAsnThrPhePheAspProSerGlyGlyGlyAspProIleLeuTyrG 232

385 GTACTTCGTT..........CCCTTTGTAGTGGTACTACTTCTT 419
    :::  ||||            ||| |||    :::::||:::||||
232 lnHisLeuPheTrpPheGlyHisProGluValTyrIleLeuIleLeu 248

420 ...............................AGCCATGTACTC 431
    |||||||:::::::
249 ProGlyPheGlyIleLeuSerHisIleVal 258
```

seq_name: SwissProt_39:COX1_MARPO

```
seq_documentation_block:
ID   COX1_MARPO     STANDARD;      PRT;    522 AA.
AC   P26856;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1 OR COXI.
OS   Marchantia polymorpha (Liverwort).
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Embryophyta; Marchantiophyta; Marchantiales;
OC   Marchantiaceae; Marchantia.
OX   NCBI_TaxID=3197;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92114051; PubMed=1731062;
RA   Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,
RA   Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;
RT   "Gene organization deduced from the complete sequence of liverwort
RT   Marchantia polymorpha mitochondrial DNA. A primitive form of plant
RT   mitochondrial genome.";
RL   J. Mol. Biol. 223:1-7(1992).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; M68929; AAC09451.1; -.
DR   PIR; S25956; S25956.
DR   HSSP; P00396; 1OCC.
DR   Mendel; 2055; MARPo;cox1;1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
```

```
FT   METAL          65      65       IRON (HEME A) (PROBABLE).
FT   METAL         244     244       COPPER B (PROBABLE).
FT   METAL         248     248       COPPER B (PROBABLE).
FT   METAL         293     293       COPPER B (PROBABLE).
FT   METAL         294     294       COPPER B (PROBABLE).
FT   METAL         379     379       IRON (HEME A3) (PROBABLE).
FT   METAL         381     381       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE      522 AA;   57551 MW;  4B88E16F03A11D01 CRC64;

alignment_scores:
         Quality:    66.00                Length:    159
           Ratio:    0.917                Gaps:        7
Percent Similarity:  45.283     Percent Identity:   22.642

alignment_block:
US-09-540-234-1 x COX1_MARPO

Align seg 1/1  to: COX1_MARPO    from: 1  to: 522

107 TGGAAACAGTCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCCTCGA 156
       ||   :  |||||||    ||:::::||::::|||.:::  :::||
107 TrpLeuLeuProProSerLeuLeuLeuLeuSerSerAlaLeuValGl 123

157 GACG.............................................. 160
       |:::|
123 uValGlyCysGlySerGlyTrpThrValTyrProProLeuSerGlyIleT 140

161 .............TCGGTGAAGCCGTAGTTGTACAAGACTAACAA 193
                  ||||| ||| :::::::::  ::|
140 hrSerHisSerGlyGlySerVal AspLeuAlaIlePheSerLeuHisLe 156

194 CGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGCTGTTTACTGTGT 243
       :::::: ::::||:::|||:::||| ::: |||||:  |||||
156 uSerGlyValSerSerIleLeuGlySerIleAsnPheIleThrThrIleP 173

244 AGCGCTTG...............ATGACTTCCTTGCTGTACTCCGCT 275
       :::::|          |||        :::
173 heAsnMetArgAlaProGlyLeuThrMetHisArgLeuProLeuPheVal 189

276 TGGAGCCATTGAGTGTCGTGTTTACCGTGCGAGCACTCGAAGACTTCCTTGC 325
       |||||| ||| ::::|||       |||:: |||  :::::
190 TrpSerValLeuValThrAlaProPheLeuLeuLeuLeuSer......LeuPr 204

326 TGTACTCGCTGGCAGCATAGTCGTGTTGGCTG............... 356
       |||||||||||:::|||::  ||||||
204 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgArgAsnPheAsnT 221

357 ............CCTGGTCGTCGGGACACCTGTGTACTACAGATACT 389
             ||| ||| |||||: |||:::
221 hrThrPhePheAspProAlaGlyGlyGlyGlyAspProIleLeuTyrGlnHis 237

390 TCCGTTT..........CCCTTGTAGTGGTACACTTCTT...... 419
       ||||            :::||||:: :::|||||||||||
238 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGl 254

420 ................AGCCATGTACTC 431
                     |||||||:::::
254 yPheGlyIleIleSerHisIleVal 262

seq_name: SwissProt_39:COX1_BETVU

seq_documentation_block:
ID  COX1_BETVU     STANDARD;      PRT;     524 AA.
AC  P24794;
DT  01-MAR-1992 (Rel. 21, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN  COX1 OR COXI.
OS  Beta vulgaris (Sugar beet).
OG  Mitochondrion.
OC  Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
```

```
OC  Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
OC  Caryophyllales; Chenopodiaceae; Beta.
OX  NCBI_TaxID=3555;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV. TK81-O;
RX  MEDLINE=91330331; PubMed=1651175;
RA  Senda M., Harada T., Mikami T., Sugiura M., Kinoshita T.;
RT  "Genomic organization and sequence analysis of the cytochrome oxidase
RT  subunit II gene from normal and male-sterile mitochondria in sugar
RT  beet.";
RL  Curr. Genet. 19:175-181(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Harada T., Mikami T., Kinoshita T.;
RT  "Nucleotide sequence of cytochrome c oxidase subunit I gene of sugar
RT  beet mitochondria.";
RL  Proc. Sugar Beet Res. 29:15-21(1987).
CC  -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC      CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC      3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC      CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC      CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC      AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC      AND COPPER B.
CC  -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC      4 FERRICYTOCHROME C.
CC  -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC  -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC      INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC  -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute.  There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; X57693; CAA40874.1; -.
DR  EMBL; M57645; AAA87330.1; -.
DR  HSSP; P00396; 1OCC.
DR  Mendel; 5481; BETVU;cox1;1.
DR  InterPro; IPR000883; -.
DR  InterPro; IPR002428; -.
DR  Pfam; PF00115; COX1; 1.
DR  PRINTS; PR01165; CYCOXIDASEI.
DR  PROSITE; PS00077; COX1; 1.
KW  Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW  Respiratory chain; Inner membrane.
FT  METAL         64      64       IRON (HEME A) (PROBABLE).
FT  METAL        243     243       COPPER B (PROBABLE).
FT  METAL        247     247       COPPER B (PROBABLE).
FT  METAL        292     292       COPPER B (PROBABLE).
FT  METAL        293     293       COPPER B (PROBABLE).
FT  METAL        378     378       IRON (HEME A3) (PROBABLE).
FT  METAL        380     380       IRON (HEME A) (PROBABLE).
FT  CONFLICT       2       9       TNLVRWLF -> AV (IN REF. 2).
FT  CONFLICT     220     226       TTFFDPA -> RPFLIRW (IN REF. 2).
FT  CONFLICT     229     229       MISSING (IN REF. 2).
FT  CONFLICT     501     503       EWM -> ND (IN REF. 2).
SQ  SEQUENCE     524 AA;    57580 MW;  A862089EE5C476EA CRC64;

alignment_scores:
         Quality:    66.00               Length:     108
           Ratio:    1.245               Gaps:         6
Percent Similarity:  49.074     Percent Identity:   27.778

alignment_block:
US-09-540-234-1 x COX1_BETVU
```

```
Align seg 1/1 to: COX1_BETVU from: 1 to: 524

    198 AATGGCGTC...GATGTTATCGGAGCCATTCGTGCTGTTACTGTGTA 244
        ::||||||    ::::::||::::|||:    ::: ||||::
    156 SerGlyValSerSerIleLeuGlySerIleAsnPheIleThrIlePh 172

    245 GCGCTTG.........ATGACTTCCTTGCTGCTACTCCGCTT 276
        :::::             ||| |||      :::
    172 eAsnMetArgGlyMetThrMetHisArgLeuProLeuPheValT 189

    277 GGAGCCATTTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTTGCT 326
        ||||| |||    |||:: |||::: |||::        |||
    189 rpSerValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203

    327 GTACTCGCTGGCAGCATAGTCTGTTTGCTG............. 356
        ||||||||||||::::  |||||
    204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220

    357 ............CCTGGTCTGGGACACCTGTGTACTACAGTACTT 390
                    |||||||
    220 rThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrGlnHisL 237

    391 CGTTT............CCCTTTGTAGTGGTACTACTTCTT....... 419
        |||        ||| |||  ::|||||||:||
    237 euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly 253

    420 ............AGCCATGTACTC 431
                ||||||:::::|
    254 PheGlyIleIleSerHisIleVal 261

seq_name: SwissProt_39:COX1_ORYSA
```

```
alignment_scores:
             Quality:    66.00           Length:     108
             Ratio:       1.245          Gaps:         6
Percent Similarity: 49.074       Percent Identity: 27.778

alignment_block:
US-09-540-234-1 x COX1_ORYSA

Align seg 1/1 to: COX1_ORYSA from: 1 to: 524

    198 AATGGCGTC...GATGTTATCGGAGCCATTCGTGCTGTTACTGTGTA 244
        ::||||||    ::::::||::::|||:    ::: ||||::
    156 SerGlyValSerSerIleLeuGlySerIleAsnPheIleThrIlePh 172

    245 GCGCTTG.........ATGACTTCCTTGCTGCTACTCCGCTT 276
        :::::             ||| |||      :::
    172 eAsnMetArgGlyMetThrMetHisArgLeuProLeuPheValT 189

    277 GGAGCCATTTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTTGCT 326
        ||||| |||    |||:: |||::: |||:::        |||
    189 rpSerValLeuValThrAlaPheLeuLeuLeuSer.....LeuPro 203

    327 GTACTCGCTGGCAGCATAGTCTGTTTGCTG............. 356
        ||||||||||||::::  |||||
    204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220

    357 ............CCTGGTCTGGGACACCTGTGTACTACAGTACTT 390
                    |||||||
    220 rThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrGlnHisL 237

    391 CGTTT............CCCTTTGTAGTGGTACTACTTCTT....... 419
        |||        ||| |||  ::|||||||:||
    237 euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly 253

    420 ............AGCCATGTACTC 431
                ||||||:::::|
    254 PheGlyIleIleSerHisIleVal 261

seq_name: SwissProt_39:COX1_ARATH
```

```
alignment_scores:
                    Quality:    66.00        Length:    108
                      Ratio:     1.245         Gaps:      6
      Percent Similarity:  49.074  Percent Identity:  27.778

alignment_block:
US-09-540-234-1 x COX1_ARATH

Align seg 1/1 to: COX1_ARATH     from: 1 to: 527

198  AATGGCGTC..GATGTATCGGAGCCATTCGTGTCGTGTTACTGTGTA  244
     ::|||||   :::::|||:::||:::||:::   :: :||:::|||
156  SerGlyValSerSerIleLeuGlySerIleAsnPheIleThrThrIlePh  172

245  GCGCTTG...........................ATGACTTCCTTGCTACTCCGCTT  276
     :::|||||                           ||:::|
172  eAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValT  189

277  GGAGCCATTTGAGTGCTGTTTACCGTGCGACCACTGCAGCACTTCCTTGCT  326
     |||||  |||        |||||||  |||:::       |||:::    |||
189  rpSerValLeuValThrAlaPheLeuLeuLeuSer......LeuPro  203

327  GTACTCGCTGGCAGCATAGTCTGTTTGCTG.......................  356
     ||||||||||:::||:::||::: ||||||
204  ValLeuAlaGlyAlaIleThrMetLeuThrAspArgAsnPheAsnTh  220

357  ........................CCTGGTCTGGGACACCCTGTACTACAGTACTT  390
                              ||||||  ||||:::    |||:::::
220  rThrPhePheAspProAlaGlyGlyGlyAspProIleLeuTyrGlnHisL  237

391  CGTTT....................CCCTTTGTAGTGGTACTACTTCTT.......  419
     |||                       |||    |||:::||::::::|||
237  euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly  253

420  ...........AGCCATGTACTC  431
              ||||||::::|:|
254  PheGlyIleIleSerHisIleVal  261
```

seq_name: SwissProt_39:COX1_OENBE

```
seq_documentation_block:
ID   COX1_OENBE     STANDARD;        PRT;    527 AA.
AC   P08743;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1 OR COI.
OS   Oenothera bertiana (Bertero's evening primrose).
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Myrtales; Onagraceae; Oenothera.
OX   NCBI_TaxID=3950;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. MUNZIA;
RA   Hiesel R., Schobel W., Schuster W., Brennicke A.;
RT   "The cytochrome oxidase subunit I and subunit III genes in Oenothera
RT   mitochondria are transcribed from identical promoter sequences.";
RL   EMBO J. 6:29-34(1987).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
```

```
seq_documentation_block:
ID   COX1_ARATH     STANDARD;        PRT;    527 AA.
AC   Q07063;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1 OR COI.
OS   Arabidopsis thaliana (Mouse-ear cress), and
OS   Raphanus sativus (Radish).
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC   Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702, 3726;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=A.thaliana; STRAIN=CV. COLUMBIA;
RX   MEDLINE=97141919; PubMed=8988169;
RA   Unseld M., Marienfeld J.R., Brandt P., Brennicke A.;
RT   "The mitochondrial genome of Arabidopsis thaliana contains 57 genes
RT   in 366,924 nucleotides.";
RL   Nat. Genet. 15:57-61(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=R.sativus;
RX   MEDLINE=91330332; PubMed=1651176;
RA   Makaroff C.A., Apel I.J., Palmer J.D.;
RT   "The role of coxI-associated repeated sequences in plant
RT   mitochondrial DNA rearrangements and radish cytoplasmic male
RT   sterility.";
RL   Curr. Genet. 19:183-190(1991).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y08502; CAA69785.1; -.
DR   EMBL; X57692; CAA40873.1; -.
DR   PIR; S14139; S14139.
DR   HSSP; P00396; 1OCC.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        64     64       IRON (HEME A) (PROBABLE).
FT   METAL       243    243       COPPER B (PROBABLE).
FT   METAL       247    247       COPPER B (PROBABLE).
FT   METAL       292    292       COPPER B (PROBABLE).
FT   METAL       293    293       COPPER B (PROBABLE).
FT   METAL       378    378       IRON (HEME A3) (PROBABLE).
FT   METAL       380    380       IRON (HEME A) (PROBABLE);
SQ   SEQUENCE   527 AA;  57996 MW;  6950995882FDE9E CRC64;
```

```
ID   COX1_MAIZE      STANDARD;      PRT;    528 AA.
AC   P08742;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1 OR COXI.
OS   Zea mays (Maize).
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC   Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Isaac P.G., Jones V.P., Leaver C.J.;
RT   "The maize cytochrome c oxidase subunit I gene: sequence, expression
RT   and rearrangement in cytoplasmic male sterile plants.";
RL   EMBO J. 4:1617-1623(1985).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X02660; CAA26496.1; -.
DR   PIR; A22840; ODZM1.
DR   HSSP; P00396; 1OCC.
DR   MaizeDB; 69214; -.
DR   Mendel; 2201; ZEAma;cox1;1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        64     64       IRON (HEME A) (PROBABLE).
FT   METAL       243    243       COPPER B (PROBABLE).
FT   METAL       247    247       COPPER B (PROBABLE).
FT   METAL       292    292       COPPER B (PROBABLE).
FT   METAL       293    293       COPPER B (PROBABLE).
FT   METAL       378    378       IRON (HEME A3) (PROBABLE).
FT   METAL       380    380       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   528 AA;  58257 MW;  75FB95E5568E89E7 CRC64;
```

alignment_scores:
```
          Quality:    66.00                     Length:        108
            Ratio:     1.245                       Gaps:          6
Percent Similarity:   49.074       Percent Identity:      27.778
```

alignment_block:
US-09-540-234-1 x COX1_MAIZE

Align seg 1/1 to: COX1_MAIZE   from: 1   to: 528

```
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X05465; CAA29025.1; -.
DR   PIR; A26170; ODOB1M.
DR   HSSP; P00396; 1OCC.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; FALSE_NEG.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL        64     64       IRON (HEME A) (PROBABLE).
FT   METAL       243    243       COPPER B (PROBABLE).
FT   METAL       247    247       COPPER B (PROBABLE).
FT   METAL       292    292       COPPER B (PROBABLE).
FT   METAL       293    293       COPPER B (PROBABLE).
FT   METAL       378    378       IRON (HEME A3) (PROBABLE).
FT   METAL       380    380       IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   527 AA;  57242 MW;  526F544B5B64ABF CRC64;
```

alignment_scores:
```
          Quality:    66.00                     Length:        109
            Ratio:     1.158                       Gaps:          8
Percent Similarity:   52.294       Percent Identity:      29.358
```

alignment_block:
US-09-540-234-1 x COX1_OENBE

Align seg 1/1 to: COX1_OENBE   from: 1   to: 527

```
198 AATGGCGTC..GATGTTATCGGAGCCATTCGTGTGTCTGTTACTCTGTA 244
    :::|||||:    ||||||||:||:|||:||:    ::: |||:::
156 SerGlyValSerIleLeuGlySerIleAsnPheIleThrThrIleSe 172

245 GCGGCTG............ATGACT.......TCCTTGCTGACTACCG 273
    ||||||            ||||||        ||||||
172 rAsnMetArgGlyLeuGlyMetThrMetHisArgSerProLeuPhe...V 188

274 CTTGGAGGCCATTGAGTGCTGTTTACCGTGCCAGCACTCGAAGACTTCCTT 323
    :::||||||  |||:|||||||:||:|||||::||:      |||
188 alTrpSerValLeuAlaThrAlaPheProIleLeuSer......Leu 202

324 GCTGTACTCGCTGGCAGCATAGTCTGTTGCTG................. 356
    |||||||||||:|||:|||:::|||||
203 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 219

357 .................CCTGGTCTGGGGACACCTGTACTACAGTA 387
    :::||||| ||| ||| ||||||||||||||  |||:::
219 nThrThrPheSerAspProAlaGlyGlyGlyAspProIleLeuTyrGlnH 236

388 ..............CCCTTTGTAGTGTGTACTACTTCTT... 419
    |||          |||:|||||:::|||
236 lsLeuPheArgPhePheGlyHisProGluValTyrIleLeuIleLeuPro 252

420 ..................AGCCATGTACTC 431
    ||||||||||:::::::
253 GlySerGlyIleIleSerHisIleVal 261
```

seq_name: SwissProt_39:COX1_MAIZE

seq_documentation_block:

```
198 AATGGCGTC...GATGTTATCGGAGCCATTCGTGTGCGTTTACTGTGTA 244
    ::||||||     :::::::|||:|||:|| :: |||:::   |||:::
156 SerGlyValSerIleLeuGlySerIleAsnPheIleThrThrIlePh 172

245 GCGCTTG...ATGACTTCCTTGCTGTGTACTCCGCTT 276
    ::::::                 |||            |||  ::::|
172 eAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValT 189

277 GGAGCCATTTGAGTGCGTGTTACCGTGCAGCACTCGAAGACTTCCTTGCT 326
    ||||| |||  ::::::::     |||         |||
189 rpSerValLeuValThrAlaPheLeuLeuLeuSer......LeuPro 203

327 GTACTCGCTGGCGAGCATAGTCGTGTTGCTG........... 356
    ||||||||||:::||: ::||||:|||:|     |||:::
204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220

357 ...........CCTGGTCTGGGACACCTGTGTACTACAGTACTT 390
    :::::::                |||:::::||         |||:::::
220 rThrPheAspProAlaGlyGlyAspProIleLeuTyrGlnHisL 237

391 CGTTT...........CCCTTTGTAGTGGTACTACTTCTT...... 419
    |||              ||| |||   :::|||:::|||
237 euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly 253

420 ...........AGCCATGTACTC 431
    ||||||||:::::|||
254 PheGlyIleIleSerHisIleVal 261

seq_name: SwissProt_39:COX1_SORBI

seq_documentation_block:
ID   COX1_SORBI   STANDARD;          PRT;     530 AA.
AC   P05502;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
GN   COX1 OR COXI.
OS   Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC   Mitochondrion.
OC   Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC   Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
OC   Andropogoneae; Sorghum.
OX   NCBI_TaxID=4558;
RN   [1]
RP   SEQUENCE OF 1-31 FROM N.A.
RX   MEDLINE=87051727; PubMed=3022938;
RA   Bailey-Serres J., Hanson D.K., Fox T.D., Leaver C.J.;
RT   "Mitochondrial genome rearrangement leads to extension and relocation
RT   of the cytochrome c oxidase subunit I gene in sorghum.";
RL   Cell 47:567-576(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Hanson D.K., Bailey-Serres J., Leaver C.J.;
RL   Submitted (MAY-1987) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 2 H(2)O +
CC       4 FERRICYTOCHROME C.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   --------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------
DR   EMBL; M14453; AAA68624.1; -.
DR   HSSP; P00396; 1OCC.
DR   Mendel; 2178; SORbi;cox1i1.
DR   InterPro; IPR000883; -.
DR   InterPro; IPR002428; -.
DR   Pfam; PF00115; COX1; 1.
DR   PRINTS; PR01165; CYCOXIDASEI.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   METAL     64     64      IRON (HEME A) (PROBABLE).
FT   METAL    243    243      COPPER B (PROBABLE).
FT   METAL    247    247      COPPER B (PROBABLE).
FT   METAL    292    292      COPPER B (PROBABLE).
FT   METAL    293    293      COPPER B (PROBABLE).
FT   METAL    378    378      IRON (HEME A3) (PROBABLE).
FT   METAL    380    380      IRON (HEME A) (PROBABLE).
SQ   SEQUENCE   530 AA;  58522 MW;  18921AE8E7689ABC CRC64;

alignment_scores:
        Quality:        66.00              Length:       108
        Ratio:           1.245             Gaps:            6
Percent Similarity:     49.074     Percent Identity:    27.778

alignment_block:
US-09-540-234-1 x COX1_SORBI

Align seg 1/1 to: COX1_SORBI    from: 1  to: 530

198 AATGGCGTC...GATGTTATCGGAGCCATTCGTGTGCGTTTACTGTGTA 244
    ::||||||     :::::::|||:|||:|| :: |||:::   |||:::
156 SerGlyValSerIleLeuGlySerIleAsnPheIleThrThrIlePh 172

245 GCGCTTG...ATGACTTCCTTGCTGTGTACTCCGCTT 276
    ::::::                 |||            |||  ::::|
172 eAsnMetArgGlyProGlyMetThrMetHisArgLeuProLeuPheValT 189

277 GGAGCCATTTGAGTGCGTGTTACCGTGCAGCACTCGAAGACTTCCTTGCT 326
    ||||| |||  ::::::::     |||         |||
189 rpSerValLeuValThrAlaPheLeuLeuLeuSer......LeuPro 203

327 GTACTCGCTGGCGAGCATAGTCGTGTTGCTG........... 356
    ||||||||||:::||: ::||||:|||:|     |||:::
204 ValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAsnTh 220

357 ...........CCTGGTCTGGGACACCTGTGTACTACAGTACTT 390
    :::::::                |||:::::||         |||:::::
220 rThrPheAspProAlaGlyGlyAspProIleLeuTyrGlnHisL 237

391 CGTTT...........CCCTTTGTAGTGGTACTACTTCTT...... 419
    |||              ||| |||   :::|||:::|||
237 euPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProGly 253

420 ...........AGCCATGTACTC 431
    ||||||||:::::|||
254 PheGlyIleIleSerHisIleVal 261

seq_name: SwissProt_39:SYR_TREPA

seq_documentation_block:
ID   SYR_TREPA    STANDARD;          PRT;     589 AA.
AC   O83803;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS).
GN   ARGS OR TP0831.
OS   Treponema pallidum.
```

```
OC   Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX   NCBI_TaxID=160;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NICHOLS;
RX   MEDLINE=98332770; PubMed=9665876;
RA   Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA   Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA   Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA   Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA   McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA   Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA   Venter J.C.;
RT   "Complete genome sequence of Treponema pallidum, the syphilis
RT   spirochete.";
RL   Science 281:375-388(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + L-ARGININE + TRNA(ARG) = AMP +
CC       PYROPHOSPHATE + L-ARGINYL-TRNA(ARG).
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AE001253; AAC65797.1; -.
DR   TIGR; TP0831; -.
DR   InterPro; IPR001278; -.
DR   InterPro; IPR001412; -.
DR   Pfam; PF00750; tRNA-synt_1d; 1.
DR   PRINTS; PR01038; TRNASYNTHARG.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT   SITE    132    142    "HIGH" REGION.
FT   SITE    386    390    "KMSKS" REGION.
SQ   SEQUENCE   589 AA;  67131 MW;  FA6156A69F4568E3 CRC64;

alignment_scores:
            Quality:  66.00           Length:     109
              Ratio:   1.200            Gaps:       5
Percent Similarity:  50.459    Percent Identity:  26.606

alignment_block:
US-09-540-234-1 x SYR_TREPA   ..

Align seg 1/1 to: SYR_TREPA  from: 1 to: 589

    1 GGGTCCACGGGCCCTTAT....................GT 20
      |||:::||||||||||||                      ::
  461 GlyAsnThrGlyGlnTyrLeuGlnTyrMetGlyAlaArgIleSerSerLe 477

   21 ATGCAGGAGCGGCCTAGTGGGTCATCTGAAGAAGGACAAAGACACCTCC 70
      : :||||:|||||||||   :: :||||||  |||||:::|||||:
  477 uLeuLysValGlnGluAspValGluGlnLysGlyProArgGluValA 494

   71 AAGCCGGACACGGCTACTGTAGCTGGCACGGCATAATGGAAAAACAGTCCC 120
      :|||: |||  ||| |||:||||  ||| ||||||||||||
  494 rgCys.AspProAlaLeuLeuThrHisGluAlaGluTrpGluLeuValLy 510

  121 TGCCCCTTTGAGGGTGCTGCTCAAACATGCCCCTC..............G 155
      ||||||  |||  ||||  :::
  510 sAlaLeuAlaArgPheProAlaCysValThrArgAlaAlaGlnGlyHisA 527

  156 AGAGCTGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCGT 205
      ::|||||| ||||| ::::::::::::
  527 spProSerValIleThrGlyThrGlyTyrLeuTyrLeuSerLysSerPheSer 543

  206 CGATGTTATCGGAGCCATTCGTCGTGCTGCTTTACTGTGTAGCGCTTGATGA 255
      ||| ||| ||||::||| :::||| ||| ||| ||| ||
  544 ArgPheTyr......HisAspCysProIleLeuCysGluAArgProAs 558

  256 CTTCCTTGCTGTACTCCGCTTGGAG 280
      :::: ||| |||||||||||||
  558 pTyr...AlaCysAlaArgLeuGlu 565

seq_name: SwissProt_39:ZG20_XENLA

seq_documentation_block:
ID   ZG20_XENLA     STANDARD;      PRT;   675 AA.
AC   P18714;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   GASTRULA ZINC FINGER PROTEIN XFG20-1 (XLCGF20.1).
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95094815; PubMed=8001572;
RA   Schaefer U., Rausch O., Bouwmeester T., Pieler T.;
RT   "Sequence-specific recognition of a repetitive DNA element by a C2H2
RT   zinc-finger protein in Xenopus.";
RL   Eur. J. Biochem. 226:567-576(1994).
RN   [2]
RP   SEQUENCE OF 85-613 FROM N.A.
RX   MEDLINE=90040698; PubMed=2509712;
RA   Nietfeld W., El-Baradi T., Mentzel H., Pieler T., Koester M.,
RA   Poeting A., Knoechel W.;
RT   "Second-order repeats in Xenopus laevis finger proteins.";
RL   J. Mol. Biol. 208:639-659(1989).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X82643; CAA57965.1; -.
DR   PIR; S06565; S06565.
DR   HSSP; P25490; 1ZNM.
DR   TRANSFAC; T02366;.
DR   InterPro; IPR000822; -.
DR   Pfam; PF00096; zf-C2H2; 18.
DR   PRINTS; PR00048; ZINCFINGER.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 18.
KW   Zinc-finger; Metal-binding; DNA-binding; Repeat.
FT   ZN_FING    62     84       C2H2-TYPE.
FT   ZN_FING    90    112       C2H2-TYPE.
FT   ZN_FING   118    140       C2H2-TYPE.
FT   ZN_FING   146    168       C2H2-TYPE.
FT   ZN_FING   174    196       C2H2-TYPE.
FT   ZN_FING   202    224       C2H2-TYPE.
FT   ZN_FING   257    279       C2H2-TYPE.
FT   ZN_FING   286    308       C2H2-TYPE.
FT   ZN_FING   344    366       C2H2-TYPE.
FT   ZN_FING   373    395       C2H2-TYPE.
FT   ZN_FING   424    446       C2H2-TYPE.
FT   ZN_FING   452    474       C2H2-TYPE.
FT   ZN_FING   507    529       C2H2-TYPE.
FT   ZN_FING   535    557       C2H2-TYPE.
FT   ZN_FING   563    585       C2H2-TYPE.
FT   ZN_FING   591    613       C2H2-TYPE.
FT   ZN_FING   642              C2H2-TYPE.
SQ   SEQUENCE   675 AA;   77116 MW;  0330948S2C1FCF39 CRC64;
```

alignment_scores:
Quality: 66.00   Length: 97
Ratio: 1.269   Gaps: 6
Percent Similarity: 53.608   Percent Identity: 26.804

alignment_block:
US-09-540-234-1/rev x ZG20_XENLA   .

Align seg 1/1   to: ZG20_XENLA   from: 1   to: 675

443 TCGAACAGACACGAGTACATGGCTAAGAAGTAGTACCAC.......... 405
     ||||||.::||||| ::::::: ::: |||
325 SerAsnGluHisGluLeuLeuThrHisGlnSerThrHisThrGluGlul 341

404 .............................TACAAAGGAAACGAA 389
                                  ::: :::::::::
341 nLysHisLeuProCysThrGluCysGlyGlyThrPheThrAsnGluGlnG 358

388 GTACTGTAGTACACAGGTGTCCCCAGACAGGCAGCAAACAGACTATGCT 339
     :: :::::: ||::   ::|||:::::|||
358 luLeuLeuAlaHisGlnSerThrHisThrGluGluGlnLysProLeuPro 374

338 GCCAGCGAGTACAGCAAGGAAGTCTTC..........GAGTGCTGCAC 301
     :::|||::| ||| :::::|| || ||
375 CysThrGluCysGly...GluIlePheSerAspGluHisGluLeuLeuTh 390

300 GGTAAACAGCACTCAAATGGCTCCAAGC........GGAGTACAG.... 264
     | :::|||||| |||||||||| |||||||
390 rHisGlnSerThrHisThrSerProSerThrGluPheGlyValGlnThrT 407

263 ..CAAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACG 225
     ::: :|||||||:: |||||||:::||||||
407 hrGluAspAsnHisGlnSer...ProSerLysAspHisThr 419

seq_name: SwissProt_39:MUCL_RAT

seq_documentation_block:
ID   MUCL_RAT   STANDARD;   PRT;   837 AA.
AC   P98089;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   INTESTINAL MUCIN-LIKE PROTEIN (MLP) (FRAGMENT).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Intestine;
RX   MEDLINE=92184794; PubMed=1371999;
RA   Xu G., Huan L.-J., Khatri I., Wang D., Bennick A., Fahim R.E.F.,
RA   Forstner G.G., Forstner J.F.;
RT   "cDNA for the carboxyl-terminal region of a rat intestinal mucin-like
RT   peptide.";
RL   J. Biol. Chem. 267:5401-5407(1992).
CC   -!- SUBUNIT: MULTIMERIC.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- TISSUE SPECIFICITY: COATS THE EPITHELIA OF THE INTESTINES.
CC   -!- SIMILARITY: CONTAINS 2 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
CC   --
CC   --
DR   EMBL; M81920; -; NOT_ANNOTATED_CDS.

DR   InterPro; IPR000359; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001007; -.
DR   InterPro; IPR001846; -.
DR   Pfam; PF00007; Cys_knot; 1.
DR   Pfam; PF00094; vwd; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01208; VWFC; 2.
DR   PROSITE; PS01185; CTCK_1; 1.
DR   PROSITE; PS01225; CTCK_2; 1.
KW   Repeat; Glycoprotein.
FT   NON_TER   1   1
FT   DOMAIN   149   837   PROBABLY IMPORTANT FOR DISULFIDE-BOND
FT   MEDIATED MUCIN POLYMERIZATION (LINK
FT   DOMAIN).
FT   DOMAIN   17   70   5 X 11 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT   17   27   1.
FT   REPEAT   28   38   2.
FT   REPEAT   39   50   3.
FT   REPEAT   51   62   4.
FT   REPEAT   63   70   5 (INCOMPLETE).
FT   DOMAIN   74   139   V (SIMILAR TO VWF A3-D4 CONJUCTION
FT   REGION).
FT   DOMAIN   140   517   I (SIMILAR TO VWF DOMAINS D1-D4).
FT   DOMAIN   472   543   VWFC 1.
FT   DOMAIN   581   648   VWFC 2.
FT   DOMAIN   732   817   CTCK.
FT   DISULFID   732   779   BY SIMILARITY.
FT   DISULFID   746   793   BY SIMILARITY.
FT   DISULFID   755   809   BY SIMILARITY.
FT   DISULFID   759   811   BY SIMILARITY.
FT   DISULFID   ?   816   BY SIMILARITY.
FT   CARBOHYD   91   91   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   164   164   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   278   278   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   289   289   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   344   344   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   410   410   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   444   444   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   515   515   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   538   538   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   612   612   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   627   627   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   695   695   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   727   727   N-LINKED (GLCNAC...   ) (POTENTIAL).
FT   CARBOHYD   749   749   N-LINKED (GLCNAC...   ) (POTENTIAL).
SQ   SEQUENCE   837 AA; 91499 MW; 6335BCDCAC897F35 CRC64;

alignment_scores:
Quality: 66.00   Length: 148
Ratio: 0.943   Gaps: 10
Percent Similarity: 47.297   Percent Identity: 23.649

alignment_block:
US-09-540-234-1 x MUCL_RAT   ..

Align seg 1/1   to: MUCL_RAT   from: 1   to: 837

32   CCTAGTGGGTCATCTGAAGGAAGGACAAAGACACCTCCAAGCCGGACAC 81
     ||||| :|:.:||||:::|||::::||
25   ProSerThrProThrProThrSerSerGlnThrThrThrProSerTh 41

82   GGCTACTGTAGCTGGCACGGCATAATGGAAAACAGTCCTGCCCTTTTGA 131
     :..:|||||:||:||::::| |||::|||:|||
41   rProSerThrThrSerSer........LysSerThrProSerThrProG 55

132 GGGTCGTCTCA............................AACATGCCCCTCGAGACGTCGGTG 166
     ::||| :||:||::::||
55   lnSerThrSerLysSerThrProSerThrProProLysThrThrLeu 71

167 AAGCCGTAGTTGTACAAGACTAACAACGCCAA..............TGGCG 204
     ::: |||

```
 72 ProGlyCysLeuAspPheAspProProArgGlnValAsnGluThrTrpThr  88
205 TCGATGTTATCGGAGCCATTCGTCGCTGTTTACTGTGTAGCGCTTGATG    254
       :::||||:::||  |||    :::|||
 88 pLeuCys..........AsnCysThrMetAlaIleCysLysTyrAspA    101
255 ACTTCCTTGCTCGTACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCG...  301
       ||:::  ::|:::|:|||||         |||  |||
101 snValValGluIleValGluLeuGlu.........CysAsnProPro    113
302 ...........TGCAGCACTCGA........AGACTTCCT.....    322
       |||||:||:             :::|||
114 ProMetProThrCysSerAsnGlyLeuLysProValArgValProAspPr  130
323 ....TGCTGTACTCGCTGCAGCAGATAGTCTGTTTGCTGCCTGGTCTG    365
       |||||  ::||||::  :::  |||  |||
130 oAspGlyCysCysTrpHisTrpGluCysAsp.CysTyrCysThrGlyTrp  146
366 GGGACACT...........GTGTACTACAGT                    386
       |||   ||      ::||||||||||
147 GlyAspProHisPheValThrPheAspGlyLeuTyrTyrSer         160


seq_name: SwissProt_39:CR2_MOUSE

seq_documentation_block:
ID   CR2_MOUSE      STANDARD;      PRT;   1025 AA.
AC   P19070;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   COMPLEMENT RECEPTOR TYPE 2 PRECURSOR (CR2) (COMPLEMENT C3D RECEPTOR).
GN   CR2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=90229735; PubMed=2139457;
RA   Fingeroth J.D.;
RT   "Comparative structure and evolution of murine CR2. The homolog of
RT   the human C3d/EBV receptor (CD21).";
RL   J. Immunol. 144:3458-3467(1990).
RN   [2]
RP   SEQUENCE OF 12-1025 FROM N.A.
RX   MEDLINE=91010789; PubMed=2145366;
RA   Molina H., Kinoshita T., Inoue K., Carel J.C., Holers V.M.;
RT   "A molecular and immunochemical characterization of mouse CR2.
RT   Evidence for a single gene model of mouse complement receptors 1 and
RT   2.";
RL   J. Immunol. 145:2974-2983(1990).
RN   [3]
RP   SEQUENCE OF 343-401 AND 991-1025 FROM N.A.
RX   MEDLINE=89098890; PubMed=2783485;
RA   Fingeroth J.D., Benedict M.A., Levy D.N., Strominger J.L.;
RT   "Identification of murine complement receptor type 2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:242-246(1989).
RN   [4]
RP   SEQUENCE OF 289-1025 FROM N.A.
RX   MEDLINE=89381350; PubMed=2528587;
RA   Kurtz C.B., Paul M.S., Aegerter M., Weis J.J., Weis J.H.;
RT   "Murine complement receptor gene family. II. Identification and
RT   characterization of the murine complement receptor type 2.";
RL   J. Immunol. 143:2058-2067(1989).
CC   -!- FUNCTION: RECEPTOR FOR COMPLEMENT C3D. PARTICIPATES IN B
CC       LYMPHOCYTES ACTIVATION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: B LYMPHOCYTES.
CC   -!- SIMILARITY: TO HUMAN C3D/EBV RECEPTOR (CD21).
CC   -!- SIMILARITY: CONTAINS 15 SUSHI (SCR) DOMAINS.
```

```
DR   EMBL; M81083; AAA37451.1; -.
DR   EMBL; M35684; AAA37448.1; -.
DR   EMBL; M61132; AAA63295.1; -.
DR   EMBL; M35685; AAA37450.1; ALT_SEQ.
DR   EMBL; M29281; AAA37447.1; -.
DR   PIR; A43526; A43526.
DR   HSSP; P10998; 1VVC.
DR   MGD; MGI:88489; Cr2.
DR   InterPro; IPR000436; -.
DR   Pfam; PF00084; sushi; 14.
KW   Complement pathway; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Receptor; Sushi.
FT   SIGNAL         1     11
FT   CHAIN         12   1025       COMPLEMENT RECEPTOR TYPE 2.
FT   DOMAIN        12    963       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     964    990       POTENTIAL.
FT   DOMAIN       991   1025       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        13     74       SUSHI 1.
FT   DOMAIN        81    139       SUSHI 2.
FT   DOMAIN       145    203       SUSHI 3.
FT   DOMAIN       206    264       SUSHI 4.
FT   DOMAIN       267    335       SUSHI 5.
FT   DOMAIN       342    399       SUSHI 6.
FT   DOMAIN       401    458       SUSHI 7.
FT   DOMAIN       462    515       SUSHI 8.
FT   DOMAIN       518    586       SUSHI 9.
FT   DOMAIN       593    649       SUSHI 10.
FT   DOMAIN       653    705       SUSHI 11.
FT   DOMAIN       708    769       SUSHI 12.
FT   DOMAIN       777    834       SUSHI 13.
FT   DOMAIN       840    898       SUSHI 14.
FT   DOMAIN       901    959       SUSHI 15.
FT   DISULFID      14     56       BY SIMILARITY.
FT   DISULFID      42     73       BY SIMILARITY.
FT   DISULFID      82    124       BY SIMILARITY.
FT   DISULFID     110    138       BY SIMILARITY.
FT   DISULFID     146    189       BY SIMILARITY.
FT   DISULFID     175    202       BY SIMILARITY.
FT   DISULFID     207    248       BY SIMILARITY.
FT   DISULFID     234    263       BY SIMILARITY.
FT   DISULFID     268    317       BY SIMILARITY.
FT   DISULFID     297    334       BY SIMILARITY.
FT   DISULFID     343    385       BY SIMILARITY.
FT   DISULFID     371    398       BY SIMILARITY.
FT   DISULFID     402    445       BY SIMILARITY.
FT   DISULFID     431    458       BY SIMILARITY.
FT   DISULFID     463    501       BY SIMILARITY.
FT   DISULFID     487    514       BY SIMILARITY.
FT   DISULFID     519    568       BY SIMILARITY.
FT   DISULFID     548    585       BY SIMILARITY.
FT   DISULFID     594    636       BY SIMILARITY.
FT   DISULFID     622    649       BY SIMILARITY.
FT   DISULFID     654    689       BY SIMILARITY.
FT   DISULFID     675    704       BY SIMILARITY.
FT   DISULFID     709    752       BY SIMILARITY.
FT   DISULFID     738    769       BY SIMILARITY.
FT   DISULFID     778    820       BY SIMILARITY.
FT   DISULFID     806    833       BY SIMILARITY.
FT   DISULFID     841    884       BY SIMILARITY.
FT   DISULFID     870    897       BY SIMILARITY.
FT   DISULFID     902    945       BY SIMILARITY.
FT   DISULFID     931    958       BY SIMILARITY.
FT   CARBOHYD      77     77       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     113    113       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     276    276       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT   CARBOHYD       316    316          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       364    364          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       380    380          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       484    484          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       527    527          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       615    615          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       639    639          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       694    694          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       754    754          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       790    790          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       813    813          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       823    823          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       851    851          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD       901    901          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CONFLICT       289    291          YGS -> EFR (IN REF. 4).
FT   CONFLICT       306    306          S -> T (IN REF. 2).
FT   CONFLICT       520    520          P -> A (IN REF. 2).
FT   CONFLICT       962    963          MISSING (IN REF. 4).
SQ   SEQUENCE      1025 AA;  112994 MW;  19E518B9A0273694 CRC64;

alignment_scores:
        Quality:     66.00              Length:      164
        Ratio:        0.917             Gaps:          9
Percent Similarity:  43.902    Percent Identity:  26.220

alignment_block:
US-09-540-234-1/rev x CR2_MOUSE    ..

Align seg 1/1  to: CR2_MOUSE  from: 1  to: 1025

441 GAACAGACACGAGACATGGCTAAGAAGAAGTACCACTACAAAGGGAAAC 392
    ||    |||  |||||||||||       :::||   :::  ||||
655 GluProMetArgValHisGly........LeuProAspSerHisIl 668

391 GAAGTACTGTAGTACACAGGTGTCCCAGACCAGGCAGCAAACAGACTAT 342
    ||   |||  ||||  |||   || :::|| |||   || |||   |||
668 eLys.LeuValLysArgThrCysGlnAsnGlyTyrGlnLeuThrGlyTyr 684

341 GCTGCCAGCGAGTACAGCAAG.................... 321
    ::  |||:::|||
685 Thr.......TyrGluLysCysGlnAsnAlaGluAsnGlyThrTrpPh 698

320 .GAAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG 272
    :  :: ::||  ||||||||    ||     |||  :::||      ||
698 eLysLysIleGluValCysThrValIleLeuCysGlnProProProLys. 714

271 GAGTACAGCAAGGAAGTCATCAAGCCGCTACACAGTAAACAGCACACGAAT 222
    ||::  :::|||::||||    :::::   :::||:::
715 .IleAlaAsnGlyHisThrGlyMetMetAlaLysHisPheLeuTyr 730

221 GGCTCC...................GATAACATCGACGCCATTGCCGT 193
    |||:::                    :::::||:::|||
731 GlyAsnGluValSerTyrGluCysAspGluGlyPheTyrLeuLeuGlyGl 747

192 TGTTAGTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTT 143
    ||||||                         ......GlnCysV 753
747 uLysSerLeu.................................

142 TGAGACGACCCTCAAAAGGGCAGGGACTGTTTTCCATTATGCCGTGCCAG 93
    :: :: |||||||||::|||   :::|||      |||:::::
753 al...AsnAspSerLysGlyHisGlySerTrpSerGlyProProProGln 768

92 CTACAGTAGCCGTGTCCGGCTTGGAGGTGTCTTTGTCCT............ 54
   |||              ::|||              |||
769 CysLeuGlnSerSerProLeuThrHisCysProAspProGluValLysHi 785

53 .......TTCCTTCAGATGACCCACTAGGCGCTCCTGCAT 21
   ||| :::   |||||  |||:::   ||:::
785 sGlyTyrLysLeuAsnLysThrHisSerAlaPheSerHis 798
```

seq_name: SwissProt_39:MA2X_HUMAN

```
seq_documentation_block:
ID   MA2X_HUMAN      STANDARD;            PRT;    1139 AA.
AC   P49641; Q13754;
DT   01-FEB-1996 (Rel. 33, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   ALPHA-MANNOSIDASE IIX (EC 3.2.1.114) (MANNOSYL-OLIGOSACCHARIDE
DE   1,3-1,6-ALPHA-MANNOSIDASE) (MAN IIX).
GN   MAN2A2 OR MANA2X.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Melanoma;
RX   MEDLINE=96102195; PubMed=8524845;
RA   Misago M., Liao Y.-F., Kudo S., Eto S., Mattei M.-G., Moremen K.W.,
RA   Fukuda M.N.;
RT   "Molecular cloning and expression of cDNAs encoding human alpha-
RT   mannosidase II and a previously unrecognized alpha-mannosidase IIx
RT   isozyme.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:11766-11770(1995).
CC   -!- FUNCTION: CATALYZES THE FIRST COMMITTED STEP IN THE BIOSYNTHESIS
CC       OF COMPLEX N-GLYCANS. IT CONTROLS CONVERSION OF HIGH MANNOSE TO
CC       COMPLEX N-GLYCANS; THE FINAL HYDROLYTIC STEP IN THE N-GLYCAN
CC       MATURATION PATHWAY.
CC   -!- CATALYTIC ACTIVITY: HYDROLYSIS OF THE TERMINAL 1,3- AND
CC       1,6-LINKED ALPHA-D-MANNOSE RESIDUES IN THE MANNOSYL-
CC       OLIGOSACCHARIDE MAN(5)(GLCNAC)(3).
CC   -!- PATHWAY: GLYCOSYLATION.
CC   -!- SUBUNIT: HOMODIMER, DISULFIDE LINKED (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI (BY
CC       SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS SEEM TO BE PRODUCED BY
CC       ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO FAMILY 38 OF GLYCOSYL HYDROLASES.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D55649; BAA09510.1; -.
DR   EMBL; L28821; AAA92022.1; -.
DR   MIM; 600988; -.
DR   InterPro; IPR000602; -.
DR   Pfam; PF01074; Glyco_hydro_38; 1.
KW   Hydrolase; Glycosidase; Transmembrane; Glycoprotein; Signal-anchor;
KW   Golgi stack; Alternative splicing.
FT   DOMAIN         1      5          CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM       6     26          SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT   DOMAIN        27    796          LUMENAL (POTENTIAL).
FT   CARBOHYD      95     95          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD     305    305          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    1082   1082          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD    1120   1120          N-LINKED (GLCNAC...)    (POTENTIAL).
FT   VARSPLIC     783    796          SIRRVDEHEQQVD -> GSGLCFLAEHPKGG (IN
FT                                     SHORT ISOFORM).
SQ   SEQUENCE    1139 AA;  129282 MW;  011CA3089FDC0028 CRC64;

alignment_scores:
        Quality:     66.00              Length:      188
        Ratio:        0.825             Gaps:         12
Percent Similarity:  42.553    Percent Identity:  27.660

alignment_block:
US-09-540-234-1 x MA2X_HUMAN    ..
```

us-09-540-234-1.rsp

Align seg 1/1   to: MA2X_HUMAN   from: 1   to: 1139

```
    9 GGGCCCTTATGTATGCAGGAGCGCCCAGTGGGTCATCTGAAGGAAAGGAC    58
      |||  ||| :::||||| ||  |||||
  716 GlyValLeuGlnLeuGlnLeuGlyLeuAspGlyHis............    727

   59 AAAGACACCTCCAAGCCGGACACGGCTA....................CTGT    90
      ::::||| |||||| ::::|||:::
  728 .ArgThrLeuProSerValArgIleTyrLeuHisGlyArgGlnLeuS   744

   91 AGCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCT   140
      :::::::|||||| |||||||||||::::
  744 erValSerArgHis........LeuArgValIleA               757

  141 CAAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAA   190
      ::::::  :::  ::::::::::      :::::||
  757 spSerGlyThrSerAspPheAlaLeuSerAsnArgTyrMetGln......   771

  191 CAACGCCAATGG........................CGTCG   207
      |||                                  ||||
  772 .....ValTrpPheSerGlyLeuThrGlyLeuLeuLysSerIleArgAr   786

  208 ATGTTATCGGAGCCATTCGTGTCGTGTTACTGTGTAGCGCTT.......   250
      |                 :::::|||||| |||        ::||
  786 gValAspGluGluHisGluGlnGlnValAspMetGlnValLeuValTyrG   803

  251 ...........GATGACTTCCTTGCTGTACTCCGCTTGGAGCCA   283
                 |||       |||  ||||| |||
  803 lyThrArgThrSerLysAspLysSerGlyAlaTyrLeuPheLeuProAsp   819

  284 TTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGCTGTACT..   331
      |||::::   ||| |||   ||   |||||||||:::::
  820 GlyGluAlaSer...ProThrSerProArgSerProProCysCysValSe   835

  332 ...........CGCTGGCAGCAT.....AGTCTGTTTG   353
                 |||||| |||  :::    ||| :::
  835 rLeuLysAlaLeuSerGlnArgTrpLeuArgThrMetSerThrPheT   852

  354 CTGCCTGGTCTGGGGACACC....................TGTGTACTA   382
      ::  |||||| |||                          |||:::
  852 hrArgSerGlyPheThrIleCysGlnGlyTrpArgGlyCysLeuTrp   868

  383 CAGTACTTCGTTCCCTTTGTAGTGGTACTACT........TCTTAGCC   423
      ||| ||||||||||                          || ||
  869 ThrTyrHisProTrpTrpTrpTrpTyrSerGlyThrThrArgSerTrpPr   885

  424 ATGTACTCGTGTC   436
      ||||||| :::
  885 oCysThrSerIle   889
```

'Mon Sep 10 07:42:23 2001   us-09-540-234-1.rpr

OM of: US-09-540-234-1 to: PIR_68:*   out_format : pfs
Date: Sep 7, 2001   5:08 PM
About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -O=/cgn2_1/USPTO_spool/US0954034/runat_07092001_145855_19433/app_query.fasta_1.507
 -DB=PIR_68 -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cd1
 -LIST=45 -DOCALIGN=200 -THR_SCORR=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US0954034 -@CGN1_1_49 -NCPU=6 -ICPU=3
 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-540-234-1
Query length: 444
Database: PIR_68:*
Database sequences: 219241
Database length: 76174552
Search time (sec): 67.920000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir2:H71287	+	85.50	157.85	0.3562	236	conserved hypothetical protein
pir2:T32076	+	85.50	149.15	0.3726	689	hypothetical protein F22E5.15 -
pir2:JH0674	+	81.50	141.73	1.04	637	L-proline transport protein - r
pir2:S59154	+	79.00	138.51	1.97	509	cytochrome-c oxidase (EC 1.9.3.
pir2:D71390	+	78.00	136.40	2.56	515	cytochrome-c oxidase (EC 1.9.3.
pir2:T08179	+	78.00	134.63	2.58	640	LRG5 protein - Chlamydomonas re
pir2:T26845	-	77.00	134.06	2.93	607	hypothetical protein Y43F4B.7
pir2:A43545	-	77.00	136.54	3.27	395	leukosialin CD43 precursor - mo
pir2:I52842	-	77.00	136.54	3.27	395	CD43 Lp-3 antigen - mouse
pir1:A47328	-	77.00	126.24	3.45	1403	natural killer cell tumor-reco
pir2:T07186	+	75.50	137.74	4.72	235	cytochrome-c oxidase (EC 1.9.3.
pir2:G71442	-	75.50	133.77	4.82	383	hypothetical protein - Arabidop
pir2:T42239	-	74.50	126.77	6.40	707	probable phosphoprotein phospha
pir2:F82619	+	74.00	129.49	7.15	447	proton glutamate symport protei
pir2:T19551	-	74.00	125.93	7.28	693	mucin-like protein let-653 prec
pir2:T02147	+	74.00	125.36	7.31	743	hypothetical protein F8K4.22 -
pir2:T09071	+	74.00	123.88	7.36	892	SH3 domains-containing protein
pir2:T28657	+	74.00	119.40	7.53	1547	blackjack protein, microtubule
pir2:G86223	+	73.50	132.05	7.99	288	hypothetical protein [imported]
pir2:T32077	+	73.50	127.94	8.16	478	hypothetical protein B24P11.150
pir2:S33642	-	73.50	113.00	8.81	3005	homeotic protein zfh-2 - fruit
pir2:D69657	-	73.00	128.38	9.21	400	methionine adenosyltransferase
pir2:S58985	+	73.00	126.36	9.31	513	cytochrome-c oxidase (EC 1.9.3.
pir2:T49859	-	73.00	126.36	9.31	513	hypothetical protein B24P11.150
pir2:S62706	+	73.00	126.20	9.32	523	cytochrome-c oxidase (EC 1.9.3.
pir2:S30291	+	72.50	125.33	10.59	514	cytochrome-c oxidase (EC 1.9.3.
pir2:S36424	+	72.50	124.96	10.61	538	cytochrome-c oxidase (EC 1.9.3.
pir2:T20654	+	72.50	121.95	10.78	779	hypothetical protein F09C8.2 -
pir2:S28066	+	72.00	126.10	11.95	413	sexual differentiation protein
pir2:T11317	-	72.00	124.18	12.06	523	cytochrome-c oxidase (EC 1.9.3.
pir2:T11646	+	72.00	121.42	12.24	735	hypothetical protein 95B7.9 - f
pir2:T25005	-	72.00	117.30	12.50	1221	probable cytochrome-c oxidase c
pir2:T36052	+	71.50	126.82	13.47	334	hypothetical protein T19C9.7 -
pir2:A44265	+	71.50	106.71	14.95	3968	trithorax homolog HTX, version
pir2:T07068	-	71.00	128.36	15.13	244	beta-fructofuranosidase (EC 3.2
pir2:A83867	-	71.00	125.62	15.35	342	homoserine dehydrogenase BH1737
pir2:D96810	-	71.00	122.06	15.63	530	hypothetical protein T1I1I1.6
pir2:T45785	-	70.50	122.60	17.65	438	hypothetical protein F26O13.180
pir2:T48357	+	70.50	118.98	17.98	684	hypothetical protein F12E4.90 -
pir2:T34433	-	70.50	115.64	18.29	1032	hypothetical protein K06A9.1a
pir2:S52957	-	70.50	112.57	18.58	1506	bimD protein - Emericella nidu

pir2:T34434   -   70.50   109.37   18.89   2232 ! hypothetical protein K06A9.
pir1:S27981   +   70.00   124.25   19.81   316 ! homoserine kinase (EC 2.7.1.
pir2:T20620   -   69.50   122.50   22.63   346 ! hypothetical protein F08H9.7
pir2:JE0132   -   69.50   119.83   22.94   481 ! glucose-1-phosphate adenylyl

seq_name: pir2:H71287

seq_documentation_block:
conserved hypothetical protein TP0731 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 26-May-2000
C;Accession: H71287
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:9832770
A;Accession: H71287
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-236 <COL>
A;Cross-references: GB:AE001245; GB:AE000520; NID:g3323033; PIDN:AAC65700.1; PID:g332
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0731
C;Superfamily: unassigned mutT domain proteins; mutT domain homology
F;113-148/Domain: mutT domain homology <MUTT>

alignment_scores:
          Quality: 85.50   Length: 128
          Ratio: 1.257   Gaps: 8
          Percent Similarity: 53.125   Percent Identity: 30.469

alignment_block:
US-09-540-234-1/rev x H71287

Align seg 1/1 to: H71287   from: 1   to: 236

392 CGAAGTACTGTA......GTACACAGTGTCCCCAGACCAGGCAGCAAAC 349
     |||||||||||| :: :|||||||| |||||||| |||
  23 ArgGluThrValGlnProAlaHisArgAlaLysGluGluLysGlnAs   39

348 AGACTATGCTGCCAGCGACTACAGCAAGGAGT..............CTTC 312
     ||||||| ||| ||| |||: ||||
  39 pAlaHisLeuAlaTrpGluValArgLysArgHisGluAlaCysArgLeuA   56

311 GAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCGGAGTACAGCA 262
     ||||||||| ||| :::::: ||||:::|| ||||:::::::::::||||
  56 rgValPheHisValGlnGluLeuGluSerValSerProArgLysThrVal 72

261 AGGAAGTCATCAAGCGCTACACAGTAAACAGCACCACGAATGGCTC...... 217
     ||||||::: ||||||::| |||||||
  73 ArgArgPheValThrLeu.........ThrAlaProGluTrpValIleva 86

216 ................CGATAACATCGA...............CGCC 201
     ::: || |||::: |||
  86 lValProHisValMetGluArgAlaGlnArgPhePheValMetValArgG 103

200 ATTGGCGTTGTTAGTCT...........TGTACAAACTAC......GGCTTC 166
     ::|||||||| ||| |||:::::::::: |||::::::::: |||
  103 lnTrpArgCysGlySerGlnThrValCysThrGluPheProGlyGlyVal 119

165 ACCGACGTCTCG...............................AGGGGCATGTTGA 140
     ||::||||| ||| |||||
  120 IleAspAlaGlyGluHisProGluAlaAlaAlaArgArgGluLeuPheGl 136

139 GACGACCCTCAAAAGGGCAGGGACTGTTTCCAT 106
     | |||::||||::::::: |||
  136 uGluThrGlyArgArgAlaSerProLeuAlaHis 147

seq_name: pir2:T32076

```
seq_documentation_block:
hypothetical protein F22E5.15 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T32076
R;Graves, T.; Wohldmann, P.; Clarke, K.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid F22E5.
A;Reference number: Z21119
A;Accession: T32076
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-689 <GRA>
A;Cross-references: EMBL:AF016681; PIDN:AAB66178.1; GSPDB:GN00020; CESP:F22E5.15
A;Experimental source: strain Bristol N2; clone F22E5
C;Genetics:
A;Gene: CESP:F22E5.15
A;Map position: 2
A;Introns: 18/3; 61/3; 134/3; 239/1; 271/1; 295/2; 359/3; 515/3; 620/1; 658/1

alignment_scores:
         Quality:   85.50                       Length:   121
         Ratio:      1.276                       Gaps:       5
Percent Similarity:  55.372     Percent Identity:   27.273

alignment_block:
US-09-540-234-1 x T32076     ..
Align seg 1/1  to: T32076  from: 1  to: 689

114 CAGTCCCTGCCC...TTTTGAGGTCGTCTCAAACATGCCCCTCGAGACG 160
    :::::||||||    :::||:::    |||
134 GluAsnLeuProGluTyrLeuSerSerPheGlnSerLeuPro........ 147

161 TCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCGTCGATG 210
    |||| |:::  |||::::::::::||:::                   :::|
148 ......AsnPheSerIleTyrGluAlaAsnSerLeuIlePheLeuValV 162

211 TTATCGGAGCCATTGCGTCGTCGTGTTACTGTGTAGCGGCCTTGATGACTTCC 260
    |||| ||:::  :::|||| :::||:::||||||||
162 aIleThrAlaValThrGlyAlaPheIleAlaPheThrPhePheThrAla 178

261 TTGCTGTACTCCGCTTGGAGCCATTGAGTGCTGTT................ 296
    ::::||||||::  ::: |||::::  :::
179 IleLeuTyrAsnIlePheArgMetLeuGlyPheMetLysLeuGlnMetSe 195

297 ...............TACCGTGCAGCACTCGAAGACTTCCTTGCTG 327
    |||||||||::||||:::||||
195 rAlaThrThrTyrLysGlnHisArgAlaAlaValTrpSerLeuIleAlaG 212

328 TACTCCGCTGGCAGCATAGTCGTTTGCTGCCTGGTCTGGGGACACCTGTG 377
    ::::||| ||||||||:::  |||            |||::
212 lnPheAlaThrSerIleIleCysPheSerPro.........ProIle 224

378 TACTACAGTACTTCGTTTCCCTTTGTAGTGGTACTACTTCTTAGCCATGT 427
    ::: |||||||||     :::  ||| ||:||:::
225 Phe..........LeuValPheValPheGlnLeuProHisAl 237

428 ACTCGTGTCTGTT 440
    ||| |||||||
237 aGlnValIleVal 241

seq_documentation_block:
L-proline transport protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 09-Oct-1992 #sequence_revision 09-Oct-1992 #text_change 28-Feb-1997
C;Accession: JH0674
R;Fremeau Jr., R.T.; Caron, M.G.; Blakely, R.D.
Neuron 8, 915-926, 1992
```

```
A;Title: Molecular cloning and expression of a high affinity l-proline transporter ex
A;Reference number: JH0674; MUID:92265305
A;Accession: JH0674
A;Molecule type: mRNA
A;Residues: 1-637 <FRE>
A;Experimental source: forebrain
C;Superfamily: gamma-aminobutyric acid transporter
C;Keywords: glycoprotein; leucine zipper; phosphoprotein; proline transport; transmem
F;46-65/Domain: transmembrane #status predicted <TM1>
F;73-93/Domain: transmembrane #status predicted <TM2>
F;76-97/Region: leucine zipper motif
F;117-137/Domain: transmembrane #status predicted <TM3>
F;217-236/Domain: transmembrane #status predicted <TM4>
F;242-264/Domain: transmembrane #status predicted <TM5>
F;291-311/Domain: transmembrane #status predicted <TM6>
F;323-345/Domain: transmembrane #status predicted <TM7>
F;374-397/Domain: transmembrane #status predicted <TM8>
F;424-443/Domain: transmembrane #status predicted <TM9>
F;457-479/Domain: transmembrane #status predicted <TM10>
F;500-519/Domain: transmembrane #status predicted <TM11>
F;537-559/Domain: transmembrane #status predicted <TM12>
F;14/Binding site: phosphate (Thr) (covalent) (by cAMP-dependent kinase) #status pred
F;41/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;182/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;239,600/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status pred

alignment_scores:
         Quality:   81.50                       Length:   105
         Ratio:      1.455                       Gaps:       2
Percent Similarity:  53.333     Percent Identity:   24.762

alignment_block:
US-09-540-234-1 x JH0674     ..
Align seg 1/1  to: JH0674  from: 1  to: 637

105 AATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCCTC 154
    :::||||||  |||:::::::||||||      :::  |||||
174 AspGlyAsnGlyAlaLeuProLeuAsnLeuSerSerThrValSerProSe 190

155 GAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCG 204
    |:::::|  |||||  :::  |||||::::::|||:::||:::||||
190 rGluGluTyrTrpSerArgTyrValLeuHisIleGlnGlySerGlnGlyI 207

205 TCGATGTTATCGGAGCCATTGCGTGCGTGTTTACTGTGTAGCGCTTGATG 254
    :::  ||||                  ::::::||||||
207 leGlyTyrArgProGlyGluIleArg................... 214

255 ACTTCCTTGCTGTACTCCGCCTTGGAGCCATTGAGTGCTGTTTACCGTGC 304
    |||:::                ::::||
215 .................TrpAsnLeuCysLeuLeuLeuLeuAl 224

305 AGCACTCGAAGACTTCCTTGCTGTACTCCGCTGGCAGCATAGTCGTTTGC 354
    :::  ||||||:::::||||||  :::
224 aTrpValIleValPheLeuCysIleLeuLeuLysGly............V 236

355 TGCCTGGTCTGGGGACACCTGTGTACTACAGTACTTCGTTTCCCTTTGTA 404
    :::  |||::||||||||:::||||||
236 aLysSerSerGlyLysValTyrPheThrAlaThrPheProTyrLeu 252

405 GTGGTACTACTTCTT 419
    :::::||||::|||
253 IleLeuLeuMetLeu 257

seq_name: pir2:S59154

seq_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - land snail mitochondrion
C;Species: mitochondrion Albinaria coerulea (land snail)
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Oct-2000
C;Accession: S59154
R;Hatzoglou, E.; Rodakis, G.C.; Lecanidou, R.
```

seq_name: pir2:JH0674

Genetics 140, 1353-1366, 1995
A:Title: Complete sequence and gene organization of the mitochondrial genome of the land
A:Reference number: S59143; MUID:96120351
A:Accession: S59154
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-509 <HAT>
A:Cross-references: EMBL:X83390; NID:g975668; PIDN:CAA58307.1; PID:g975680; GSPDB:GN0013
C:Genetics:
A:Gene: COI
A:Genome: mitochondrion
A:Genetic code: SGC4
A:Start codon: TTG
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
transmembrane protein
F:1-8/Domain: mitochondrial matrix #status predicted <MM1>
F:8-452/Domain: cytochrome-c oxidase chain I homology <CO1>
F:9-37/Domain: transmembrane #status predicted <TM01>
F:38-46/Domain: intracristal #status predicted <ITC1>
F:47-82/Domain: transmembrane #status predicted <TM02>
F:83-90/Domain: mitochondrial matrix #status predicted <MM2>
F:91-113/Domain: transmembrane #status predicted <TM03>
F:114-136/Domain: intracristal #status predicted <ITC2>
F:137-166/Domain: mitochondrial matrix #status predicted <TM04>
F:167-178/Domain: transmembrane #status predicted <MM3>
F:179-208/Domain: mitochondrial matrix #status predicted <TM05>
F:209-223/Domain: intracristal #status predicted <ITC3>
F:224-256/Domain: transmembrane #status predicted <TM06>
F:257-264/Domain: intracristal #status predicted <ITC1>
F:265-281/Domain: transmembrane #status predicted <TM07>
F:282-293/Domain: intracristal #status predicted <ITC4>
F:294-322/Domain: transmembrane #status predicted <TM08>
F:323-330/Domain: mitochondrial matrix #status predicted <MM5>
F:331-351/Domain: transmembrane #status predicted <TM09>
F:352-365/Domain: intracristal #status predicted <ITC5>
F:366-395/Domain: transmembrane #status predicted <TM10>
F:396-401/Domain: mitochondrial matrix #status predicted <MM6>
F:402-428/Domain: transmembrane #status predicted <TM11>
F:429-441/Domain: intracristal #status predicted <ITC6>
F:442-473/Domain: transmembrane #status predicted <TM12>
F:57,373/Binding site: heme a iron (His) (axial ligands) #status predicted
F:236,285,286/Binding site: copper (His) #status predicted
F:236-240/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:240/Binding site: oxygen (Tyr) #status predicted
F:363/Binding site: magnesium (His) (shared with chain II) #status predicted
F:371/Binding site: heme a3 iron (His) (axial ligand) #status predicted

alignment_scores:
Quality: 79.00   Length: 143
Ratio: 1.097   Gaps: 8
Percent Similarity: 50.350   Percent Identity: 27.972

alignment_block:
US-09-540-234-1 x S59154
Align seg 1/1 to: S59154 from: 1 to: 509

```
 92 GCTGGCACGGCATAATGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    |||||||||::: |||  :: || |||  ::  |||
118 AlaGlyThrGly...TrpPheValTyrPro.ProLeu......SerSerS 131

142 AAACATGGCCGTCGATGTTATCGGACGCCATTCGTCGTGTTACTGT 191
    :: ::: |||  :: ::::::::||: ::
131 erLeuAlaHisSerGlyAlaSerValAspLeuAlaIlePheSerLeuHis 147

192 AACGCCAATGGCGTCGATGTTATCGGAGCCGTAGTTGTGTTACTGT 241
    ||| |||  :::  :||||||||||:::: :: :::::::||: ::
148 LeuAlaGlyMetSerSerIleLeuGlyAlaIleAsnPheIleThrThrIl 164

242 GTAGCGCTTG.................ATGACTTCCTTGCTGCTACTCCG 273
    : ::::::                  ||| :
```

164 ePheAsnMetArgSerProGlyMetThrMetGluArgValSerLeuPheV 181

```
274 CTTGGAGCCATTTGAGTGCTGTTACCGTGCACTGCAGACTTCCTT 323
    ::|||||| ||| |||| ::|||::: |||::  |||:::||
181 alTrpSerIleLeuValThrValPheLeuLeuLeuLeuSer.....Leu 195

324 GCTGTACTCGCTGGCAGCAGTCGTTGCTG................ 356
    |||||||||||||:::|||:  ||||||
196 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgArgAsnPheAs 212

357 ...........CCTGGTCTGGGGACACCTGTGTACTACAGTA 387
               ||| ||| |||:: |||:::  |||::
212 nThrSerPheAspProAlaGlyGlyAspProIleLeuTyrGlnH 229

388 CTTCGTTT...........CCCTTTGAGTGCTGTACTACTTCTT.... 419
    |||       ||| :::||||::::|||
229 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 245

420 ....................AGCCATGTACTC 431
                        ||||||::::|||
246 GlyPheGlyMetIleSerHisIleLeu 254
```

seg_name=: pir2:D71390

seg_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - common lancelet mitochondrion
C:Species: mitochondrion Branchiostoma lanceolatum (common lancelet)
C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: D71390
R:Spruyt, N.; Delarbre, C.; Gachelin, G.; Laudet, V.
Nucleic Acids Res. 26, 3279-3285, 1998
A:Title: Complete sequence of the amphioxus (Branchiostoma lanceolatum) mitochondrial
A:Reference number: A71390; MUID:98292550
A:Accession: D71390
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-515 <SPR>
A:Cross-references: GB:Y16474; NID:g3292989; PIDN:CAA76249.1; PID:g3292993
C:Genetics:
A:Gene: COI
A:Genome: mitochondrion
A:Genetic code: SGC4
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane
transmembrane protein
F:11-457/Domain: cytochrome-c oxidase chain I homology <CO1>
F:61,378/Binding site: heme a iron (His) (axial ligands) #status predicted
F:239,289,290/Binding site: copper (His) #status predicted
F:239-243/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:243/Binding site: oxygen (Tyr) #status predicted
F:368/Binding site: magnesium (His) (shared with chain II) #status predicted
F:376/Binding site: heme a3 iron (His) (axial ligand) #status predicted

alignment_scores:
Quality: 78.00   Length: 115
Ratio: 1.300   Gaps: 5
Percent Similarity: 52.174   Percent Identity: 25.217

alignment_block:
US-09-540-234-1 x D71390

Align seg 1/1 to: D71390 from: 1 to: 515

```
174 AGTTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCGGACGCCAT 223
    :::::::::||||||||||::
146 AlaIlePheSerLeuHisLeuAlaGlyValSerSerIleLeuGlyAlaIl 162

224 TCGTCGTGTTACTGTTACTGTGTAGGCGTCGATGACTTCCTTGCTGCTACTCC.. 272
    ||| :: :: ||| :::: :::::: ::::::
162 eAsnPheIleThrIleHisAsnMetArgAlaSerIleGluTrpAsnA 179

273 ...................GCTTGGGAGCCATTTGAGTGCTGTTACCGTGCAGCA 308
    : ::::
```

179 rgValProLeuPheValTrpSerIleTrpValThrAlaTyrLeuLeuLeu 195
    ::::||||||
309 CTCGAAGACTTCCTTGCTGTACTCGCTGGCAGCACATAGTCTGTTGCTG.. 356
    |||.:::: ||| ||||||||||.::|||:.|||:.| ||||||
196 LeuSer.....LeuProValLeuAlaGlyAlaIleThrMetLeuLeuTh 210

357 ...............................CCTGGCTGGGGACAC 372
                                    ||| ||| |||||||||
210 rAspArgAsnIleAsnThrPhePheAspProSerGlyGlyGlyAspP 227

373 CTGTGTACTACAGTACTTCGTT....................CCCTTTGTAGTG 407
    |||.:: |||:::| || |||
227 rOIleLeuTyrGluHisLeuPheTrpPhePheGlyHisProGluValTyr 243

408 GTACTACTTCTT......................AGCCATGTACTCGTG 434
    ::||:|:.:|||
244 IleLeuIleLeuProGlyPheGlyIleIleSerHisIleIleIle 258

seq_name: pir2:T08179

seq_documentation_block:
LRG5 protein - Chlamydomonas reinhardtii
C;Species: Chlamydomonas reinhardtii
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C;Accession: T08179
R;Gloeckner, G.; Beck, C.F.
submitted to the EMBL Data Library, October 1996
A;Description: Molecular characterization of a gene (LRG5) involved in blue light signal
A;Reference number: Z16399
A;Accession: T08179
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-640 <GLO>
A;Cross-references: EMBL:U73817; NID:g1644369; PID:g1644370
C;Genetics:
A;Gene: LRG5

alignment_scores:
          Quality: 78.00   Length: 192
          Ratio: 1.099   Gaps: 8
Percent Similarity: 36.979   Percent Identity: 21.875

alignment_block:
US-09-540-234-1 x T08179   ..

Align seg 1/1   to: T08179 from: 1 to: 640

11 GCCCTTATGTATGCAGGAGCCGCCTAGTGGGTCATCTGAAGGAAAGGACAA 60
   ||||||:||: ||||||||| |||||| ||||::| |||||:| ||
427 AlaAlaValProAlaAlaAlaCysSerGlyPheArgGlyGlyArgGlyGl 443

61 AGACACCTCCAAGCCGGACACGGCTACTGTAGCTGGCACGCATAATGGA 110
   :::::|||||||||| |||||| ||| ||||:: |||||||||||
443 yGlyValAlaArg.........ValAlaAlaGlyThrSerArgAlaA 456

111 AAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACG 160
   |||||| :||||| ||||||||||||
456 laGlyValProArgLeu........................ 462

161 TCGGTGAAGCCCAGTTGTTGTACAAGACTAACAACGGCACCGCCAATGGCGTCGA... 208
   |||||| ||||||||| |||: |||||||||||
463 ....GlnArgArgTrpArgArgAr 469

208 208
   ||||| ||||||| |||| ||||||
469 gGlyArgGlyTrpArgArgValArgArgArgArgGlyValAlaGlyArgA 486

209 ....TGTTATCGGAGCCATTCGTGTCGTGTTACTGTAGCGGCTGAT 253
    ||| |||||| ||: ||| |:||
486 laValCysThrAlaGlyLysArgCysCysTrpMetThrCysLeuProMetTrp 502

seq_name: pir2:T26845

seq_documentation_block:
hypothetical protein Y43F4B.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26845
R;Matthews, L.
submitted to the EMBL Data Library, January 1998
A;Reference number: 220276
A;Accession: T26845
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-607 <WIL>
A;Cross-references: EMBL:AL021481; PIDN:CAA16336.1; CESP:Y43F4B.7
A;Experimental source: clone Y43F4B
C;Genetics:
A;Gene: CESP:Y43F4B.7
A;Introns: 53/3; 80/3; 110/2; 141/2; 169/3; 229/1; 261/3; 300/2; 320/2; 355/3; 377/1;

alignment_scores:
          Quality: 77.50   Length: 136
          Ratio: 1.062   Gaps: 8
Percent Similarity: 53.676   Percent Identity: 24.265

alignment_block:
US-09-540-234-1 x T26845   ..

Align seg 1/1   to: T26845 from: 1 to: 607

58 CAAAGACACCTCCAAGCCGGACACGGCTACTGTAGCTGGCACGCATAAT 107
   ||||| :|||||:||||:: |||||| |||||| ||||||:|||||
266 GluAsnArgMetGlnSerProHisAlaPheIleSerTrpAsnGlyValLe 282

108 GGAAAAAGTCCCTGC.........................CCTT 127
   :|||||| ||| ||| |||
282 u...AsnSerSerCysLeuValValLeuAlaIlePheSerValThrGlyIyP 298

128 TTGAGGGTCGTCTCAAACATGCCGTCGATGTTATCGGAGCCTAGTT 177
   : ||||||||:::   :|||||| ||||: ||: ||||||:
298 heTyrGlyTyrLeuSer...LeuLeuProSer...LeuAsnAspThrAlaThr 313

178 TGTACAAGACTAACAACGGCACCGCCAATGGCGTCGATGTTATCGGAGCCATTCGT 227
   ||||||::: |||||| ||||||| ||| ||||||::::||
314 LeuAsnLeuProMetThrProPheTyrGlnThrIleLysLeuMetPheVa 330

228 GTGCTGTTTACTGTGTAGCG......CTTGATGACTTCCTTGCTGTACTC 271
   |:::||||: |||| :||| |||:: :::
228 laValCysThrAlaGlyLysArgCysCysTrpMetThrCysLeuProMetTrp 502

254 GACTTCCTTGCTGTACTCGGCTTGGAGCCATTGAG.................... 289
    :::|||||: |||||:
503 GlySerGlyThrTrpProTrpProArgProLeuMetThrProSerArgTh 519

290 ....TGCTGTTACCGTGCAGCACTCGAAGACTTCCTTGCTGTACT 331
    :::|||||||||| |||:::|||:::
519 rCysAlaCysLeuPro.........ThrProCysCysSerArgT 531

332 ..................................CGCTGGCAGCAT 343
                                      ||||||||:::|
531 rpLeuArgArgTrpArgCysGlyTrpAlaProGlyGlyArgTrpArgCys 547

344 AGTCTGTTTGCTCCCTGG..........TCTGGGGACACCTGTGTACT 381
    ||||| :::::||| ||| ||||| ||| :||
548 SerLeuCysSerCysTrpArgTrpGlyCysSerGlyArgThrProLeuLe 564

382 ACAGTACTTCGTT...........TCCCTTTGTAGTGGTACTA 413
    ||||:|:|||:: |||:::|||||||
564 uProThrTrpTrpValTrpArgArgArgCysCysArgCysCysArgGlySerA 581

414 CTTCTTAGCCATGTACTCGTGTCTGT 439
    ||||| :::|||||||
581 rgAlaProArgCysAsnTrpValCys 589

```
330 lAlaCysIleMet.IleSerTyrProLeuGlnPheTyrValProMetGlu 346

272 CGCTTGGAGCCATTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCC 321
    |||::||| ||   ::
347 ArgIleGluLysTrp................IleThrArgLysIlePr 357

322 TTGCTGTACTCGCTGGCAGCATAGTCTGTTTGCCTGCCTGGTCTGGGGAC. 370
    |||   :::   |||||| ::||||||
357 oValAspLysGlnThrLeuTyrIleTyrIleAlaArgTyrSerGlyVal 374

371 ....ACCTGTGTA..............CTACAGTACTTCGTT 394
    ||||::|||   |||   |||||||
374 leLeuThrCysAlaIleAlaGluLeuIleProHisLeuAlaLeuPheIle 390

395 TCCCTT 400
    ||||||
391 SerLeu 392

seq_name: pir2:A43545

seq_documentation_block:
leukosialin CD43 precursor - mouse
N;Alternate names: large sialoglycoprotein; sialophorin; T-cell surface glycoprotein CD4
C;Species: Mus musculus (house mouse)
C;Date: 06-Nov-1992 #sequence_revision 06-Nov-1992 #text_change 23-Jul-1999
C;Accession: A43545; S12702; A45842; S08065
R;Cyster, J.; Somoza, C.; Killeen, N.; Williams, A.F.
Eur.J. Immunol. 20, 875-881, 1990
A;Title: Protein sequence and gene structure for mouse leukosialin (CD43), a T lymphocyt
A;Reference number: A43545; MUID:90269342
A;Accession: A43545
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-395 <CYS>
A;Cross-references: GB:X17018; NID:g52891; PIDN:CAA34884.1; PID:g52892
R;Dorfman, K.S.; Litaker, W.; Baecher, C.M.; Frelinger, J.G.
Nucleic Acids Res. 18, 4932, 1990
A;Title: The nucleotide sequence of Ly 48 (mouse leukosialin, sialophorin): the mouse ho
A;Accession: S12702
A;Molecule type: DNA
A;Residues: 1-395 <DOR>
A;Cross-references: EMBL:X52609; NID:g55500; PIDN:CAA36840.1; PID:g55501
R;Baecher, C.M.; Dorfman, K.S.; Mattei, M.G.; Frelinger, J.G.
Immunogenetics 31, 307-314, 1990
A;Title: cDNA cloning and localization of the mouse leukosialin gene (Ly48) to chromosom
A;Reference number: A45842; MUID:90316596
A;Accession: A45842
A;Molecule type: mRNA
A;Residues: 345-383 <BAE>
A;Cross-references: GB:M30693; NID:g198912; PIDN:AAA39457.1; PID:g198913
C;Genetics:
A;Gene: Ly48
A;Map position: 7
C;Superfamily: leukosialin
C;Keywords: glycoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-395/Product: leukosialin #status predicted <MAT>

alignment_scores:
            Quality:  77.00        Length:        140
            Ratio:   1.132         Gaps:            7
Percent Similarity: 48.571  Percent Identity: 30.714

alignment_block:
US-09-540-234-1/rev x A43545    ..

Align seg 1/1  to: A43545  from: 1   to: 395

412 AGTACCACTACAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCCCCAG 363
    |||:||| |||   ||||||||   :::||||
179 SerSerLysThrSerGlyProProValThrThrAlaThrGlySerLeuGl 195
```

```
362 ACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAGG...... 320
    |||:||||||   ||||||||||||||||:::||||||
195 yProSerGluMetHisGlyLeuProAlaThrAlaThrSerV 212

319 .AAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG 272
    :::||||||||   |||:|||   |||:::
212 alGluSerSerValAlaArgGlyThrSer............... 222

271 GAGTACAGCAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGAAT 222
    ||| ||:||| :::||||   :::|||||
223 ......ValSerSerArgLysThrSerThrThrSerThrGl 234

221 GGCTCCGATAACATCG..ACGCCA........ 200
    :  ||||||||||| ::||
234 nAspProIleThrThrArgSerProSerGlnGluSerSerGlyMetLeuL 251

199 ..................TTGGCGTTG 191
                      :::||||||
251 euValProMetLeuIleAlaLeuValLValLeuAlaLeuValAlaLeu 267

190 TTAGTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTG 141
    |||::||:::|||   |||   |||||   :::|||
268 LeuLeuLeuTrpArgGlnArgGlnLysArgArg.ThrGlyAlaLeu.... 282

140 AGACGACCCTC................AAAAGGGCAGGGACTGTTTTCCATTAT 103
    ||||||| |||   |||||||||||||::|||
283 ......ThrLeuSerGlyGlyLysArgAsnGlyValValAlaProAlaTrp 297

102 GCCGTGCCAGCTACAGTA 85
    |||  ||||||  |||
298 AlaGlyProAlaArgVal 303

seq_name: pir2:I52842

seq_documentation_block:
CD43 Lp-3 antigen - mouse
C;Species: Mus sp. (mouse)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 23-Jul-1999
C;Accession: I52842
R;Shiota, J.; Nishimura, H.; Okamoto, H.; Yu, B.; Hattori, S.; Abe, M.; Okada, T.; No
Cell. Immunol. 155, 402-413, 1994
A;Title: A unique murine CD43 epitope Lp-3: distinct distribution from another CD43 e
A;Reference number: I52842; MUID:94236703
A;Accession: I52842
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-395 <RES>
A;Cross-references: GB:S70677; NID:g547746; PIDN:AAB30765.1; PID:g546747
C;Superfamily: leukosialin

alignment_scores:
            Quality:  77.00        Length:        140
            Ratio:   1.132         Gaps:            7
Percent Similarity: 48.571  Percent Identity: 30.714

alignment_block:
US-09-540-234-1/rev x I52842    ..

Align seg 1/1  to: I52842  from: 1   to: 395

412 AGTACCACTACAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCCCCAG 363
    |||:||| |||   ||||||||   :::||||
179 SerSerLysThrSerGlyProProValThrThrAlaThrGlySerLeuGl 195

362 ACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAGG...... 320
    |||:||||||   ||||||||||||||||:::||||||
195 yProSerGluMetHisGlyLeuProAlaThrAlaThrSerV 212

319 .AAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG 272
    :::||||||||   |||:|||   |||:::
212 alGluSerSerValAlaArgGlyThrSer............... 222
```

```
271  GAGTAGCAGCAAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGAAT  222
          :::|||    ||    :::||||
223  ..........................ValSerSerArgLysThrSerThrThrSerThrGl  234

221  GGCTCCGATAACATCG...ACGCCA..............................  200
     : ||||||||:::      :::|||
234  nAspProIleThrThrArgSerProSerGlnGluSerSerGlyMetLeuL  251

199  ......................................................TTGGCCGTTG  191
                                                            ::|||||||
251  euValProMetLeuIleAlaLeuValValValLeuValAlaLeuAlaLeu  267

190  TTAGTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTG  141
     |||:::||.:|  ||   ::::::
268  LeuLeuLeuTrpArgGlnArgGlnLysArgArg.ThrGlyAlaLeu....  282

140  AGACGACCCTC.............AAAAGGGCAGGGACTGTGTTTCCATTAT  103
     ||||||                  |||:::|||      :::|||
283  ....ThrLeuSerGlyGlyGlyLysArgAsnGlyValValAspAlaTrp  297

102  GCCGTGCCAGCTACAGTA  85
     ||| |||:||   |||
298  AlaGlyProAlaArgVal  303
```

seq_name: pir1:A47328

seq_documentation_block:
natural killer cell tumor-recognition protein - human
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-tumo
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 24-Sep-1999
C:Accession: A47328
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cells.
A:Reference number: A47328; MUID:93133824
A:Accession: A47328
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1403 <AND>
A:Cross-references: GB:L04288; NID:g181251; PIDN:AAA35734.1; PID:g181252
A:Experimental source: NK killer cells from adult blood
C:Genetics:
A:Gene: GDB:NKTR
A:Cross-references: GDB:137171; OMIM:161565
A:Map position: 3p23-3p21
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYP>

```
alignment_scores:
      Quality:    77.00           Length:   76
      Ratio:       1.638          Gaps:      3
Percent Similarity: 61.842    Percent Identity: 28.947
```

alignment_block:
US-09-540-234-1/rev x A47328

Align seg 1/1 to: A47328 from: 1 to: 1403 ..

```
437  AGACACCAGTACATGGCTAAGAAGTAGTACCAC.........TACAAAGG  397
     :::||:::     :::|||   |||               |||||
826  LysHisSerSerSerGluLysThrLeuHisSerLysTyrValLysGl  842

396  GAAA...CGAAGTACTCTAGTACACAGGTGTCCCCAGACCAGGCAGCAAA  350
     |:::   ::::::||:::|||:::|
842  yArgAspArgSerCysValArgLysTyrSerGluSerArgSerSerL  859

349  CAGACTATGCTGCCAGCGAGTACAGCAGCAAGGAAGTCTTCGAGTGCTGCACG  300
     |||:::||:::||:::::::
```

```
859  euAspTyrSerSerAspSer......................GluGlnSerSer  869
          :|||:::|||:||               :::
299  GTAAACAGCACCATCCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCA  250

870  ValGlnAlaThrGlnSerAlaArgLysGluLysGluLysGluLysGlnGlnMetGl  886
     |||:::|||:::|||   |||          :::|||||||
249  AGCGCTACACAGTAAACAGCACACGAAT  222

886  uArgThrHisAsnLysGlnLysLysAsn  895
       :   |||:::|||||||||:::
```

seq_name: pir2:T07186

seq_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - Chlorella vulgaris mitochondrion (fragmen
C:Species: mitochondrion Chlorella vulgaris
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: T07186
R:Watanabe, K.I.; Ehara, M.; Inagaki, Y.; Ohama, T.
Gene 213, 1-7, 1998
A:Title: Distinctive origins of group I introns found in the COXI genes of three gree
A:Reference number: Z15982; MUID:98372069
A:Accession: T07186
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-235 <WAT>
A:Cross-references: EMBL:AB011523; NID:g2943728; PIDN:BAA25069.1; PID:g2943729
A:Experimental source: strain NIES-227
C:Genetics:
A:Gene: COXI
A:Genome: mitochondrion
A:Note: intron positions not resolved (incomplete sequence)
C:Function:
A:Description: catalyzes the oxidation of four molecules of reduced cytochrome c in t
producing two protons
A:Pathway: oxidative phosphorylation; respiratory chain
C:Superfamily: cytochrome c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; magnesium; membrane-assoc
F:109,158,159/Binding site: copper (His) #status predicted
F:109-113/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:113/Binding site: oxygen (Tyr) #status predicted

```
alignment_scores:
      Quality:    75.50           Length:   115
      Ratio:       1.280          Gaps:       5
Percent Similarity: 51.304    Percent Identity: 26.957
```

alignment_block:
US-09-540-234-1 x T07186

Align seg 1/1 to: T07186 from: 1 to: 235 ..

```
174  AGTTTGTACAAGACTAACAACGCCAATGCGTCGATGTGTATCGGAGCCAT  223
     :::     |||||    :::|||||||      :::||||||
15   AlaIlePheSerLeuHisLeuAlaGlyAlaSerSerIleMetGlyAlaIl  31

224  TCGTGTGCTGTTTACTGTGTAGCGCTTG..................ATGA  255
     |:::::|||||:::::|||:::              ||
31   eAsnLeuIleThrIlePheAsnMetArgAlaProGlyMetSerMetH  48

256  CTTCCTTGCTGCTACTCCGCTTGGAGCCATTTGAGTGCTGTTTACCGTGCA  305
     :::|||||| |||   :::|||||||   ||   :::|||
48   isArgLeuProLeuPheValTrpSerValLeuIleThrAlaPheLeuLeu  64

306  GCACTCGAAGACTTCCTTGCTGCTACTCCGCTGGAGCCATAGTCTGTTTGCT  355
     ::::|||:::    ||:::|||:::|||:::::||
65   IleLeuSer.....LeuProValLeuAlaGlyLyIleThrMetLeuLeu  79

356  G.................................CCTGGCTGGGGA  369
     |||                               :::
79   uThrAspArgAsnPheAsnThrPheLeuAspProAlaGlyGlyGlyA  96

370  CACCTGTGTACTACAGTACTTCGTT................CCCTTTGTA  404
     |||
```

```
                                   |||::::        |||:::         |||     |||
                          96  spProIleLeuTyrGlnHisLeuPheTrpPhePheGlyHisProGluVal 112

                         405  GTGGTACTACTTCTT...............AGCCATGTACTC 431
                                  ||||||||::|       |||||||||::|
                         113  TyrIleLeuIleLeuProGlyPheGlyIleValSerHisValIle 127

                   seq_name: pir2:G71442

                   seq_documentation_block:
                   hypothetical protein - Arabidopsis thaliana
                   C;Species: Arabidopsis thaliana (mouse-ear cress)
                   A;Variety: columbia
                   C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
                   A;Accession: G71442
                   R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirk
                   P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giel
                   avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
                   erhoft, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans
                   Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
                   Nature 391, 485-488, 1998
                   C.; Chalwatzis, N.
                   A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
                   A;Reference number: A71400; MUID:98121113
                   A;Accession: G71442
                   A;Status: preliminary; nucleic acid sequence not shown; translation not shown
                   A;Molecule type: DNA
                   A;Residues: 1-383 <BEV>
                   A;Cross-references: GB:Z97343; NID:g2245073; PID:g2245094
                   C;Genetics:
                   A;Map position: 4COP9-4G3845

                   alignment_scores:
                          Quality:     75.50          Length:    110
                            Ratio:      1.198          Gaps:        6
                   Percent Similarity:  57.273    Percent Identity:  27.273

                   alignment_block:
                   US-09-540-234-1/rev x G71442   ..

                   Align seg 1/1  to: G71442  from: 1 to: 383

                         402 CAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCC................ 367
                             |||::::|||  :::|||||  ||| |||      ::||||::||||||
                          53 GlnInGlnAspGlnPheCysValGlyGlnAsnSerProProValLeuPh  69

                         366 .CCAGACCAGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAA.... 322
                             ||||||  |||||:::|     ||| :::|||||||:::::
                          69 eProAspAlaAlaAlaAspProLeuValAlaGlyLyLysIleMetAsnSerV  86

                         321 ..GGAAGTCTTCGAGTG.................CTGCACGGTAAA    295
                             :||::|||  ||||:::|||||               ||||||||||
                          86 aIGlyThrArgArgAlaGlyThrLeuSerLysTrpPheHisHisAsnLys 102

                         294 CAGCACTCAAATGGCTCCAGCGAGTACAGGAGGAAGTCATCAAGCGC   245
                             ::|||||||::::||||  :::::   ||| |||||||:::|||||||
                         103 GluHisSerThrSerSerThrThrAsnLeuLysLysLysAspLysSerVa 119

                         244 TACACAGTAAACACCACACGAATGGCTCCGATAACATCGACGCCATTGGC 195
                             ::||||||           :::::     ||| :::|||||||||:::
                         119 lArgValGluAsnAlaHisVal.........His.SerAlaValSer     131

                         194 GTTGTTAGTCTTGTACAAACTACGGCTTCACCGGACGTCTCCGAGGGCATG 145
                             ||||| |||||| |||||:::     ||| |||||||||  ::: :|||
                         132 IleAlaAlaLeuAlaLaGlyLeuAlaSerValThrSerAlaSerAsnCy 148

                         144 TTTGAGACGACCCTCAAAAGGGCAGGGA 117
                             |         :::||| |||||||::||
                         148 s...........LysGlySerGly 152

                   seq_name: pir2:T42239
```

```
                   seq_documentation_block:
                   probable phosphoprotein phosphatase (EC 3.1.3.16) - Caenorhabditis elegans
                   C;Species: Caenorhabditis elegans
                   C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
                   C;Accession: T42239
                   R;Sherman, P.M.; Sun, H.; Macke, J.P.; Williams, J.; Smallwood, P.M.; Nathans, J.
                   Proc. Natl. Acad. Sci. U.S.A. 94, 11639-11644, 1997
                   A;Title: Identification and characterization of a conserved family of protein serine/
                   A;Reference number: Z22115; MUID:97471020
                   A;Accession: T42239
                   A;Status: preliminary; translated from GB/EMBL/DDBJ
                   A;Molecule type: mRNA
                   A;Residues: 1-707 <SHE>
                   A;Cross-references: EMBL:AF023454; NID:g2586408; PIDN:AAB82794.1; PID:g2586409
                   C;Keywords: EF hand; phosphoric monoester hydrolase; serine/threonine-specific phosph

                   alignment_scores:
                          Quality:     74.50          Length:    192
                            Ratio:      0.776          Gaps:        9
                   Percent Similarity:  50.000    Percent Identity:  23.958

                   alignment_block:
                   US-09-540-234-1/rev x T42239

                   Align seg 1/1  to: T42239  from: 1 to: 707    ..

                         442 CGAACAGACACCGAGTACATGGCTAAGAAGTAGT.............AC 408
                             |||  ::    :||||| |||:::|||
                           9 ArgGlnAsnProSerThrGluLeuLysSerThrArgAlaThrArgTh  25

                         407 CACTACAAAGGGAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAG 358
                             ||||||||| ||||:::|||||              |||:::
                          25 rThrThrSerSerSerGlnArgAsnAsnTyrAsn..............A 37

                         357 GCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGT 308
                             :::::||::|||   :::|||||||      :::|||:::|||||||
                          37 spAsnAsnGlnAsnThrSerSerSerGlyAsnLysLysGluSerSer    53

                         307 GCTGCACGGTAAACACCACTCAAAT..GGCTCCAAGCGGGAGTACAGCAA 261
                             ||||||::|||||||||||::|       ||||||:::||:::|||||
                          54 SerSer.SerLysGlnHisSerSerLysSerLysLysSerAsnSerL   70

                         260 GGAAGTCATCAAGCGGCTACA......CAGTAAACA....... 232
                             :||||:::  |||  |||          ||||||||
                          70 ysLysAsnArgSerProSerProGlnLeuThrIleLysSerAla  86

                         231 .....GCAACGAATGGCTCCGA..................TA 212
                             ::::::|||||:::|||  |||
                          87 IleLeuIleGlnLysTrpTyrArgArgCysGluAlaArgLeuGluAlaAr 103

                         211 ACATCGACGCCATTGGCGTTGTTAGTCTTGTTACAAACTACGGC..... 169
                             ::||| |||| |||::::|||  :::|||
                         103 gArgArgAlaThrProGlnIlePheThrAlaLeuGluTyrAlaGlyGluG 120

                         168 ............................TTCACCGGACGTCTCGAGGGGCATG 145
                             ::|||:::|||||:::|||||  :::|||
                         120 lnAspGlnLeuLysLeuTyrAspPhePheAlaAspValIleArgAlaMet 136

                         144 TTTGAGACGACCCTCAAAAGGGCA................ 121
                             ||| |:::     :::
                         137 AlaGluGluAsnGlyLysGlyGlyValGluAsnGlyArgAsnSerProLe 153

                         120 ...GGGACTGTTTCCATTATGCCGTGCCAGCTACAGTAGCCGTGTCCG 75
                             ::::::     |||||||||  |||::: :||      :::||
                         153 uMetSerAlaSerHisTyrAlaLysProSerLeuMetAspSerGluG 170

                          74 GCTTGGAGGTGTCTTTGTCCTTTCCTTCAGATGACCCACTAGGCGCTCCT 25
                             ||   |||        |||:::     ||||||:::|||:::|||||||
                         170 lyGluThrValLysMetLeuGluAspThrSerProThrAsnValAsp 186

                          24 GCATACATAAGGGCCCGTGGACCC 1
                             :::  ::|||||||
```

187 IleAspArgAsnTyrLysGlyPro 194

seq_name: pir2:F82619

seq_documentation_block:
proton glutamate symport protein xF1937 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: F82619
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: F82619
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-447 <SIM>
A;Cross-references: GB:AE004013; GB:AE003849; NID:g9107030; PIDN:AAF84739.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Carraro, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
C;Contents: annotation
C;Genetics:
C;Gene: XF1937
C;Superfamily: C4-dicarboxylate carrier protein

alignment_scores:
     Quality: 74.00   Length: 104
     Ratio: 1.321   Gaps: 3
Percent Similarity: 53.846   Percent Identity: 26.923

alignment_block:
US-09-540-234-1 x F82619

Align seg 1/1 to: F82619 from: 1 to: 447

113 ACAGTCCTGCCCTTTTGAGGGTCGTCTCAAACATGCCCCTCGAGACGTC 162
    |||.||||:||| |||||||||||||||||||||:||| || |
311 ThrLeuProThrAlaLeuArgValAlaAspGluMetLysLeuProLysGl 327

163 GGTGAAGCCGT.........AGTTTGTACAAGACTAACAACGCCAATGGC 203
    :::|| |||||| ::||||| ||| ||| ||||||||||||
327 nIle.SerArgPheValLeuThrIleGlyAlaThrAlaAsnGlnAsnGly 343

204 GTCGATGTTATCGGAGCCATTCGTGTCGTTTACTCTGTAGCGCTTGAT 253
    ||||:|||| |||||||| :::|| ||| |||:|| || ||| :::
344 ThrAlaLeuPheGluGlyLeuThrValIlePheLeuLeuAlaGlnPhePheGl 360

254 GACTTCCTTGCTGTACTCCGCTTGGACCCATTTGAGTGCTGTTTACCGTG 303
    :::|||||| || ||||| :::||| ::: |||
360 yValAspLeuSerPheSerGlnGlnAlaMetValMetAla.......... 373

304 CAGCACTCGAAGACTTCCTTGCTGTACTCGCTGGCAGCATAGTCTGTTTG 353
    |||||:|||||:|||||:   :::|||||||
374 ....ValCysIle 376

354 CTGCCTGGTCTGGGGACACCTGTACTACAGTACTTCGTTTCCCTTTGT 403
    ||| |||:|||||| || |||:|||||| ||||| |||:||||
377 LeuGlyLeuGlyIleGlyThrAlaGlyValProSerGlySerLeuProValVa 393

404 AGTGGTACTA 413
    |:||:||:::
393 lAlaLeuIle 396

seq_name: pir2:T19551

seq_documentation_block:
mucin-like protein let-653 precursor - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C;Accession: T19551; S69552; S59520
R;Dobson, R.
submitted to the EMBL Data Library, May 1996
A;Reference number: Z19141
A;Accession: T19551
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-693 <WIL>
A;Cross-references: EMBL:Z72504; PIDN:CAA96602.1; GSPDB:GN00022; CESP:C29E6.1
A;Experimental source: clone C29E6
R;Jones, S.J.M.
submitted to the EMBL Data Library, February 1996
A;Reference number: S69552
A;Accession: S69552
A;Molecule type: DNA
A;Residues: 1-693 <JON>
A;Cross-references: EMBL:X91045; NID:g1199510; PID:e222516; PID:g1199511
A;Note: this is a revision to the sequence from reference S59520
R;Jones, S.J.M.; Baillie, D.L.
Mol. Gen. Genet. 248, 719-726, 1995
A;Title: Characterization of the let-653 gene in Caenorhabditis elegans.
A;Reference number: S59520; MUID:96069711
A;Accession: S59520
A;Molecule type: DNA
A;Residues: 1-40, 'T', 42-138, 'PQVSY', 144, 'MDNNSKPRRVSFELFGWK', 162-317,'I',319-693 <JON>
A;Cross-references: EMBL:X91045
A;Note: this sequence has been revised in reference S69552
C;Genetics:
A;Gene: let-653
A;Map position: 4
A;Introns: 21/1; 119/1; 212/1; 280/3; 334/2; 374/2; 473/2; 596/3; 658/1
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-693/Product: mucin-like protein let-653 #status predicted <MAT>

alignment_scores:
     Quality: 74.00   Length: 151
     Ratio: 1.104   Gaps: 6
Percent Similarity: 44.371   Percent Identity: 24.503

alignment_block:
US-09-540-234-1/rev x T19551

Align seg 1/1 to: T19551 from: 1 to: 693

439 ACAGACACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGA 390
    |||:|||||| ||||||:|||||::||||:|||||||||||||||||
407 ThrGluThrAlaThrSerSerSerThrThrThrValThr........... 420

389 AGTACTGTAGTACACAGGTGTCCCAGCAGGCAGCAAACAGACTATGC 340
    :::|| |||| ||| :::||| ||| ||| ||| |||
421 ....ThrGlnLysProThrThrValThrSerThrThrThrL 433

339 TGCCAGCGAGTACAGCAAGGAGTCTTCGAGTGCTGCACGGTAAACAGCA 290
    ||||:|||||:||||||||| :::||||||
433 euProSerThrThrAlaSerThrThrThrLys.............. 443

289 CTCAAATGCTCCAAGCGAGTACAGCAAGGAGTACAAGAAGTCATCAAGCGCTACAC 240
    |||||||||:   :::||| :::||||| |||
444 ....Th 444

239 AGTAAACAGCACACGAATGGCTCCGATAACATCGACG............ 203
    |:::||| ||| |||::|||
444 .....ThrHis

Left column:

```
444  rThrThrSerThrProThrSerProGlnThrThrThrThrHisValGlyA  461
          ::: :   ||||  |     :|  :|: : : |::
202  ..CCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCACCGACGTCTC  155

461  laProAlaSerValAlaSerVal.......AlaHis.AspGlySe  474
          || :     ||| |||  ||
154  GAGGGGCATGTTTGAGACGACCCTCAAAAGGGCAGGACTGTTTTCCATT  105
                :::                  ::|||||
474  rThrLeuAlaGlyLysProLysValProValIlePheAspIlePheHisA  491

104  ATGCCGTGCCAGCTACAGTAGCCCGTGCCGGCTTGGAGTGTCTTTGTCC  55
          :::  :|||::  |||: | ||| |||   :  :::|||::|||
491  snGlyGlnProValGluAlaValAlaValGlyThrLysIleSerLeuSer  507

54   TTTCCTTCAGATGACCCCACTAGGCGCTCCTGCATACATAAGGGCCGTGG  5
          |||                ||| |||:  |:|||:
508  PheArgProHisTyrProIle...ProProGluTyrValAspValArgGl  523

4    A  4
          |
523  y  523
```

seq_name: pir2:T02147

seq_documentation_block:
hypothetical protein F8K4.22 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 22-Oct-1999
C;Accession: T02147
R;Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Kwan, A.; Yu, G.; Oji, O.; Liu, S.; Li,
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome 1 BAC F8K4 sequence.
A;Reference number: Z14574
A;Accession: T02147
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-743 <VYS>
A;Cross-references: EMBL:AC004392; NID:g3282170; PID:g3367535; GSPDB:GN00059; ATSP:F8K4.
C;Genetics:
A;Gene: ATSP:F8K4.22
A;Map position: 1

alignment_scores:
          Quality: 74.00   Length: 132
          Ratio: 1.156   Gaps: 8
Percent Similarity: 48.485   Percent Identity: 23.485

alignment_block:
US-09-540-234-1 x T02147   ..

Align seg 1/1 to: T02147 from: 1 to: 743

```
97   CACGGCATAATGGAAAAACAGTCCCGCCCTTTGAGGGTCGT.......  138
          |||::::::::      ||| ||||||||::|||:   ::
496  HisSerValAspProCysSerAlaCysProTrpLysCysLysAlaGlyPh  512

139  ...CTCAAACATGCCCCTCGACGTCGGTGAA...............G  169
          ::                   ||
512  ePheTyrGluCysGlnGlnAspLysAspLysGluTyrLeuLysLysLeuH  529

170  CCGTAGTTTGTACAAGACTAACAACGCCAATGGCGTCGATGTTATCGGAG  219
          || |||:: |||::::: |||:::::: ||
529  isValGlnCysAlaThrMetSerGluProLeuValHisLysSerHisGlu  545

220  CCATTCG..............................TGTGCTG  233
          ||||||                            ||||:|
546  .HisSerLeuPheLeuThrSerLysProGluGlnGlnArgArgCysSerV  562

234  TTTACTGTGTAGGCGCTTGATGACTTCCTTGCTGTACTCCGCTTGGAGCCA  283
          :|          ||         ::  :|::|||
562  al...CysLysLysLeuGluGlnPheSerThrAsnGluThrPheAsnCys  577
```

Right column:

```
284  TTTGAGTGC.............................TGTTTACCGTGCAGCACTCGAAGACT  318
          :::||||||                        |||::|||  |   |   ||
578  IleGluCysAspTyrPheAlaLeuCysPheAlaLeuCysGlyCysAlaThr.......Le  592

319  TCCTTGCTGTACTCGCTGGCAGCATAGT..........CTGTTGTGCTGCCT  359
          ||||        :::|||:::|:|||:||           :::|||
592  uProGlnGluValArgTyrGluHisAspLysHisMetLeuThrLeuSerT  609

360  GGTCTGGGACACCTGTACTACACAGTACTTCGTTTCCTTTGT  403
          :::::   :::|||||||||   ||| :|||| :::::|||
609  yrGlyGluThrCysIleThrThrTyrTrpCysAspIleCys  623
```

seq_name: pir2:T09071

seq_documentation_block:
SH3 domains-containing protein POSH -. mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Sep-2000
C;Accession: T09071
R;Tapon, N.; Nagata, K.; Lamarche, N.; Hall, A.
EMBO J. 17, 1395-1404, 1998
A;Reference number: Z16552; MUID:98151363
A;Accession: T09071
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-892 <TAP>
A;Cross-references: EMBL:AF030131; NID:g3002587; PIDN:AAC40070.1; PID:g3002588
A;Experimental source: cell line Ras-transformed NIH 3T3 cells
A;Note: activates JNK/SAPK cascade; Rac-binding protein
C;Genetics:
A;Gene: POSH
C;Superfamily: RING finger homology
C;Keywords: signal transduction
F;8-58/Domain: RING finger homology <RRN>

alignment_scores:
          Quality: 74.00   Length: 125
          Ratio: 1.104   Gaps: 7
Percent Similarity: 53.600   Percent Identity: 28.000

alignment_block:
US-09-540-234-1/rev x T09071

Align seg 1/1 to: T09071 from: 1 to: 892

```
398  GGGAAACGAAGTACTGTAGTACACAGGTGTCCCAGACCAGGCAGCAAAC  349
          |||:::: ||||||||||::||| ||||||||||||||   :::::
440  GlySerSerGluGlnIleAlaHisLeuArgProGlnThrArgProSerVa  456

348  AGACTATGCTGCCAGCCAGTACAGC.....AAGGA............  318
          |||      ||:::|:   |||   |||||
456  lTyrValAlaIleTyrProTyrProArgLysGluAspGluLeuGluL  473

317  ...............................GTCTTCGAGTGCTGCACG.........  300
          ::::::|||||| ||||
473  euArgLysGlyMetPheLeuValPheGluArgProGlnThrArgProTrp  489

299  GTAAACAGCACTCAAATGGCTCCAAGC.......GGAGTACAGCAAGGAAG  256
          ::::::||:|| :|||     |||||::::|||  |||||
490  TyrLysGlyThrSerMetHisThrSerLysIleGlyValPheProGlyAs  506

255  TCATCAAGCGCTACACAGTAAACAGCACAGCACGAATGGCTCCGAT......  213
          :::   ||    ::||:| |||||::::|:|||
506  nTyrValAlaProValThrArgAlaValThrAsnAlaSerGlnAlaLysV  523

212  ..AACATCGACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGGCTTCA  165
          ::::|| |||||::|| |: ||| |||::|||  :::|||:|||
523  alSerMetSerThrAlaGlyGlnAlaSerArgGlyValThrMetValSer  539

164  CCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGGCAGGGACT  115
          |||:||||| |||     :::||||:::|||| ::|||||||||
```

540 ProSerThrAlaGlyGlyProThrGlnLysPro...GlnGlyAsnGlyVa 555
114 GTTT.......TCCATTATGCCG 99
          |||:::::||||
555 lAlaGlyAsnProSerValValPro 563
seq_name: pir2:T28657

seq_documentation_block:
blackjack protein, microtubule associated - American bird grasshopper
C;Species: Schistocerca americana (American bird grasshopper)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-May-2000
C;Accession: T28657
R;Zachow, K.R.; Bentley, D.
J. Cell Sci. 109, 1497-1507, 1996
A;Title: Blackjack, a novel protein associated with microtubules in embryonic neurons.
A;Reference number: Z20498; MUID:9393048
A;Accession: T28657
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1547 <ZAC>
A;Cross-references: EMBL:L76606; NID:g1220510; PID:g1220511; PIDN:AAC37266.1
A;Experimental source: embryo; clone library Kai Zinn

alignment_scores:
        Quality: 74.00   Length: 131
        Ratio: 1.072   Gaps: 7
Percent Similarity: 52.672   Percent Identity: 30.534

alignment_block:
US-09-540-234-1/rev x T28657

Align seg 1/1 to: T28657 from: 1 to: 1547

439 ACAGACACGGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGA 390
    |||::::   ::::||||||||||||||||
1326 ThrSerLeuAlaThrGlnLysGlnSerThrThrThr.....GlyLy 1340

389 AGTACTGTAGTACACAGGTGTCCCAGCAGCAGCAAA.......... 350
    :::   :::::||:::||:::||||||
1340 sValProThrSerIleAlaProLysProSerSerLysThrAlaLeuS 1357

349 ..CAGATCATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCG......... 311
   :::|||| |||:::||:::||::::||:::||||
1357 erSerThrThrLysProSerGlnThrValArgAlaAlaSerAlaProLys 1373

310 ......AGTGCTGCACGGTAAACAGCCACTCAAA.....TGGCT 279
       ||||||| ||||| |||
1374 ProArgValProSerAlaAlaThrThrAlaThrLysAsnProLeuAl 1390

278 CCAAGCGGAGTACAGCAAGGAAGTCATCAAGCGCTACACAGTAAACAGCA 229
    :::||||:::   :::|| |:::||
1390 aAsnGlyAspValLysSerAlaGluValLysLysThrSerValAlaThrS 1407

228 CACGAATGGCTCCGATAAACATCGACGCCATTGGCGTTGTTAGTCTTGTAC 179
    :::|||||||| ||| |||||||
1407 erLysIleAlaProArgThrHisThrAlaProSer............ 1418

178 AAACTACGGCTTCACCGACGTCTCGAGGGGCCATGTTGAGACGACCCTCA 129
    ||| ||| |||||
1419 ......ArgThr.ThrThrValSer............L 1425

128 AAAGGGCAGGACTGTTTCCATTATGCCGTGCCAGCTACA 88
    :::|||||||:::   ::|||||||| |||
1425 ysSerAlaGlyThrThrThrAlaAlaSerValProLysThr 1438

seq_name: pir2:G86223

seq_documentation_block:
hypothetical protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C;Accession: G86223
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Davis, R.W.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: G86223
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-288 <STO>
A;Cross-references: GB:AE005172; NID:g3249107; PIDN:AAC24090.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1

alignment_scores:
        Quality: 73.50   Length: 141
        Ratio: 1.131   Gaps: 6
Percent Similarity: 46.099   Percent Identity: 24.113

alignment_block:
US-09-540-234-1/rev x G86223

Align seg 1/1 to: G86223 from: 1 to: 288

435 ACAGAGTACATGGCTAAGAAGTAGTACCACTACAAAGG.......... 397
    :::||:::|||||||| ::::||||||:::|||
162 SerArgAlaTyrGlyLeuAspLeuValProAlaGluThrSerIleLysVa 178

396 ..........GAAACGAAGTACTGTAGTACACAGGTGTCCCCAGACCAGG 357
              || || |||||||||||||
178 lGlyAsnGlyGluLysIleLysSerThrTyrLeuSer........... 191

356 CAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTG 307
    ||| ::::: :::|||| |||||
192 ....CysLeuAspAsnLysLysGlnGlnMetGluArgVal 203

306 CTTGCACGTAAACAGCACTCAAATGGCTCAAGCGGTACAGCAAGGAA 257
    :::::||:::||| |||
204 PheTyrGlyGln............ArgGl 209

256 GTCATCAAGCGCTACACAGTAAACAGCACGACAGCACGATGGCTCCGATAAACATC 207
    ::: ||||| :::|||| ::::|||
209 uGlnArgMetAlaThr.............HisGluValArgSerHisA 222

206 GACGCCATTGGCGTTGTTAGTCTTGTACAAACTACGCGTTCACCGACGTC 157
    ||||| :::   ::::||| :::::||||||
222 rgArgGlu.........ProGluValArgAspAspLysThrSerValAlaIle 235

156 TCGAGGGCATGTTTGAGACG.......... 136
    ::: || ||||:::||||
236 GluLeuGlyGluPheGluThrGlySerGlyGluGluGlyGlyAspAspLysGly 252

135 .ACCCTCAAAAGGCAGGGACGGACTGTTTTCCATTATGCCGTGCCAGCTACAGG 87
    ::::: || :::|||| :::|||
252 uValValMetSerLeuThrGluValLysValLysGlyTyrGlnLeuLysGlyGlyI 269

86 TAGCCGTGTCCGGCTTGGAGGTG 64
   :::|||||||:::||||||
269 leAlaIleAspGlyIleGluVal 276

seq_name: pir2:T32077

seq_documentation_block:
hypothetical protein F22E5.16 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C;Accession: T32077
R;Graves, T.; Wohldmann, P.; Clarke, K.
submitted to the EMBL Data Library, July 1997
A;Description: The sequence of C. elegans cosmid F22E5.
A;Reference number: Z21119
A;Accession: T32077
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-478 <GRA>
A;Cross-references: EMBL:AF016681; PIDN:AAB66179.1; GSPDB:GN00020; CESP:F22E5
A;Experimental source: strain Bristol N2; clone F22E5
C;Genetics:
A;Gene: CESP:F22E5.16
A;Map position: 2
A;Introns: 30/2; 182/1; 241/3; 314/3; 419/1; 458/2

alignment_scores:
   Quality: 73.50   Length: 122
   Ratio: 1.131   Gaps: 5
Percent Similarity: 53.279   Percent Identity: 27.049

alignment_block:
US-09-540-234-1 x T32077

Align seg 1/1 to: T32077 from: 1 to: 478

111 AAACAGTCCCTGCCC...TTTTGAGGGTCGTCTCAAACATGCCCCTCGAG 157
    ::::::::|||||| ::: :|||| |||:::|| |||
313 GluGluAsnLeuProGluTyrLeuSerSerPheGlnSerLeuPro..... 327

158 ACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAATGGCGTCG 207
    ::: :||:::|||| : :: |||:::||::|| :::||
328 ....AsnPheSerIleTyrGlnAlaAsnAlaLeuLeuPheIleM 341

208 ATGTTATCGGAGCCATT.................... 224
    |||||| ||:::
341 etValIleThrAlaValIleGlyGlyLeuLeuAlaPheSerPhePheMet 357

225 CGTGTGCTGTTTACTGTGTGAGCCGTTGATGACTTCCTTGCTGTACTCCGC 274
    ||||||::::::: :|| ::::: ::::::: :::
358 AlaValLeuTyrAsnIlePheArgMetLeuSerPheMetLysValGlnMe 374

275 TTGGAGCCATTTGAGTGGTTTACCGTGCAGCACTCGAAGACTTCCTTG 324
    ::: ::::::: ::::: :::::::|| ::::||:::
374 tSerAspThrThrTyrLysArgHisArgAlaAlaValTrpSerLeuIleA 391

325 CTGTACTCGCTGGCAGCATAGTCGTGTTTGCTGCCTGGTCTGGGGACACCT 374
    || :::|||| ||||||||:::||:::|||||| |||
391 laGlnPheAlaThrSerIleIleCysPheLeuPro............Pro 403

375 GTGTACTACAGATACTTCGTTCCCTTTGTAGTGTACTACTTCTTAGCCA 424
    ::: ||||||| ||||||||| ||| ||| ::
404 Ile..........SerLeuValPheValValPheLeuLysLeuProAs 416

425 TGTACTCGTCTGTT 440
   : ||| |||
416 nProGlnValIleVal 421

seq_name: pir2:S33642

C;Accession: T32077
A;Cross-references: EMBL:M63450; NID:g158822; PIDN:AAA29051.1; PID:g158823
C;Genetics:
A;Gene: zfh-2
A;Cross-references: FlyBase:FBgn0004607
C;Superfamily: unassigned homeobox proteins; homeobox homology
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation; zinc finger
F;1798-1854/Domain: homeobox homology <HOX1>
F;2155-2211/Domain: homeobox homology <HOX2>
F;2761-2817/Domain: homeobox homology <HOX3>

alignment_scores:
   Quality: 73.50   Length: 121
   Ratio: 1.185   Gaps: 3
Percent Similarity: 51.240   Percent Identity: 21.488

alignment_block:
US-09-540-234-1/rev x S33642

Align seg 1/1 to: S33642 from: 1 to: 3005

410 TACCACTACAAAGGAAACGAAGAAGTACTGTAGTACACAGG........... 372
    :::|||||||||| :::::::|| :::||
566 TrpHisTyrLysTyrGlnGluThrLeuGluIleHisMetArgGluLysHi 582

371 ......................................TGTCCCCAGACCAGGCAGCAAA 350
                     ||| :::||||:::
582 sProAspGlyGluSerAlaCysGlyTyrCysLeuAlaGlyGlnGlnHisP 599

349 CAGACTATGCTGCCAGCGAG...............TACAGCAAGGAAGTC 315
    ||| :|||| :::|| :::
599 roArgLeuAlaArgGlyGluSerTyrSerCysGlyTyrLysProTyrArg 615

314 TTCGAGTGCTGCACCGTAAACAGCACTCAAATGGCTCCAAGCGGAGTACA 265
    ||||||| :::::   :::|| :::||
616 CysGluIleCysAsnTyrSerThrThrThrLysGlyAsnLeuSerIleHi 632

264 GCAAGGAGGTCATCAAGCGGCTACACAGTAAACAGCACAGAATGGCTCCG 215
    ::: ||| ::::::: |||:::||| :::|||:::
632 sMetGlnSerAspLysHisLeuAsnAsnMetGlnGlnLeuAsnSerSerG 649

214 ATAACATCGACGCCATTGGCGTTGTTAGT................CTTGTA 180
    :::|||::: ||| :::||| |||:::
649 lnAsnMetValAlaAlaAlaAlaAlaAlaAlaValThrGlyLysLeuLeu 665

179 CAACTACGGCTTCACCGACGTCTCGAGGGCATGTTTGAGACGACCCCTC 130
    :::::||||||||||||::: :::::|||| :|||||||
666 LeuSerSerSerProGlnValThrAlaAlaCysProSerAsnSerGl 682

129 AAAAGGGCAGGGA 117
   :::||| |||
682 ySerGlyAlaGly 686

seq_name: pir2:D69657

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: D69657
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-400 <KUN>
A;Cross-references: GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB15033.1; PID:g2635539
A;Experimental source: strain 168
C;Genetics:
A;Gene: metK
C;Superfamily: methionine adenosyltransferase
C;Keywords: S-adenosylmethionine; transferase

alignment_scores:
            Quality: 73.00   Length: 57
              Ratio: 1.825   Gaps: 0
Percent Similarity: 70.175   Percent Identity: 26.316

alignment_block:
US-09-540-234-1/rev x D69657   ..

Align seg 1/1 to: D69657 from: 1 to: 400

365 CAGACCAGGCAGCCAAACAGACTATGCTGCCAGCGAGTACAGCCAAGGAAGT 316
      |||||| ::::::: |||: ||||::||| :::||:
 73 GlnThrIleLysGluIleGlyTyrThrArgAlaLysTyrGlyPheAspAl 89

315 CTTCGAGTGCTGCACCGTAAACAGCACTCAAATGGCTCCAAGCGGGAGTAC 266
      |||||||||||||||||||| ::: ::: ||:   :::
 89 aGluThrCysAlaValLeuThrSerIleAspGluGlnSerAlaAspIleA 106

265 AGCAAGGAAGTCATCAAGCCTACACAGTAAACAGCACACACGAATGGCTCC 216
      :::|| |||||||||||||::: ||||||||||||||||
106 laMetGlyValAspGlnAlaLeuGluAlaArgGluGlyThrMetSerAsp 122

215 GATAACATCGACGCCATTGGC 195
      ||:::::||:::|||||||||
123 GluGluIleGluAlaIleGly 129

seq_name: pir2:S58985

seq_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - earthworm (Lumbricus terrestris) mitochondri
C;Species: mitochondrion Lumbricus terrestris (common earthworm)
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 07-Dec-1999
A;Accession: S58985
R;Boore, J.L.; Brown, W.M.
Genetics 141, 305-319, 1995
A;Title: Complete sequence of the mitochondrial DNA of the annelid worm Lumbricus terres
A;Reference number: S58985; MUID:96042914
A;Accession: S58985
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-513 <BOO>
A;Cross-references: EMBL:U24570; NID:g984290; PIDN:AAC46864.1; PID:g984291
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC4
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
F;8-454/Domain: cytochrome-c oxidase chain I homology <CO1>
F;58,375/Binding site: heme a iron (His) (axial ligands) #status predicted
F;237,287,288/Binding site: copper (His) #status predicted
F;237-241/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F;241/Binding site: oxygen (Tyr) #status predicted
F;365/Binding site: magnesium (His) (shared with chain II) #status predicted
F;373/Binding site: heme a3 iron (His) (axial ligand) #status predicted

---

alignment_scores:
            Quality: 73.00   Length: 159
              Ratio: 1.028   Gaps: 7
Percent Similarity: 44.654   Percent Identity: 24.528

alignment_block:
US-09-540-234-1 x S58985   ..

Align seg 1/1 to: S58985 from: 1 to: 513

107 TGGAAAACAGTCCTGCCCTTTTGAGGGTCGTCTCAAACATGCCCCTCGA 156
      ||| ::::::|||:::||:|| :::|||||| ::|||||||
100 TrpLeuLeuProSerLeuIleLeuLeuLeuValSerSerAlaAlaValGl 116

157 G.. 157
    |
116 uLysGlyAlaGlyThrGlyTrpThrValTyrProProLeuAlaSerAsnL 133

158 ...............ACGTCGGTGAAGCCGTAGTTGTACAAGACTAACAA 193
                   ||||| :::   :::||||||:::
133 euAlaHisAlaGlyProSerVal.AspLeuAlaIlePheSerLeuHisLe 149

194 CGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTCGTGTTTACTGTGT 243
      :::: |||:||||||||| ||:||| ||||||||| ::: ||||||
149 uAlaGlyAlaSerSerIleLeuGlyAlaIleAsnPheIleThrThrVal 166

244 AGCGCTTGATGACTTCCTTGCTG.................TACTCCGCT 275
      :::|| ||||::|| |||:|||||
166 leAsnMetArgTrpSerGlyLeuArgLeuArgIleProLeuPheVal 182

276 TGGAGCCATTGAGTGCTGTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
      ||:::| ||| :::|| ||:|||||||||::: |||
183 TrpAlaValLeuIleThrValValLeuLeuLeuSer.....LeuPr 197

326 TGTACTGCTGGCAGCATAGTCTGTTTGCTG.................... 356
      |||||||||||||:|| ::::|||:||||
197 oValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnLeuAsnT 214

357 ...................CCTGCTGCGGGACACCTGTCTACTACAGTACT 389
                         |||:: ||| |||:::|||| :::
214 hrSerPheAspProAlaGlyGlyAspProIleLeuTyrGlnHis 230

390 TCGTTT....................CCCTTTGTAGTGGTACTACTTCTT..... 419
    ||| ||| ||| :::||:::|||
231 LeuPheTrpPheGlyHisProGluValTyrIleLeuIleLeuProGl 247

420 .............................AGCCATGTACTC 431
                                 ||||||||||||
247 yPheGlyAlaIleSerHisIleVal 255

seq_name: pir2:T49859

seq_documentation_block:
hypothetical protein B24P11.150 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 18-Aug-2000
C;Accession: T49859
R;Schulte, U.; Algn, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu
submitted to the Protein Sequence Database, May 2000
A;Reference number: 225022
A;Accession: T49859
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-513 <SCH>
A;Cross-references: EMBL:AL356833; GSPDB:GN00116; NCSP:B24P11.150
A;Experimental source: BAC clone B24P11; strain OR74A
C;Genetics:
A;Gene: NCSP:B24P11.150
A;Map position: 6
A;Introns: 81/3; 380/2
C;Superfamily: Neurospora crassa hypothetical protein B24P11.150

alignment_scores:
Quality: 73.00   Length: 116
Ratio: 1.281   Gaps: 4
Percent Similarity: 49.138   Percent Identity: 25.862

alignment_block:
US-09-540-234-1/rev x T49859   ..

Align seg 1/1 to: T49859 from: 1 to: 513

364 AGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAGGAAGTC 315
     :::|||||||:::   :::   ::|||||||:::::||| |||::|
414 GlnProGlyThrSerGlySerProMetProSerThrThrProArgGluAr 430

314 TTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAGCG........ 272
     ::: ||| ||| |||::::|||
430 gArgSerLeuArgIleLeuAspLeuHisArgLeuArgAlaAlaSerI 447

271 ..GAGTACAGCAAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGA 224
     ||| |||:::::||::::||| :::::::   :::
447 leGluGluArgMetGluIleLeuArgArgHisArgSerGlnGlnGlnGln 463

223 ATGGCTCCGATAACATCGACGCCATTGGCGTTGTTAGTCTTGTACAAACT 174
     :::||:::::||| |||||:::
464 ArgArgHisValThrGlySerSerLeuThr.................... 473

173 ACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGG 124
                                        ||||||:::::
474 ....GlyAspProAspThrG 479

474 ....GlyAspProAspThrG 479
                                                    |||||||:::::
123 GCAGGGACTGTGTTTCCATTATGCCGTGCCAGCTACAGTAGCCGTGTCCGG 74
     ||::::::::||| ||||||| :::::||| |||
479 lyGluSerAlaAspProThrHisArgAlaArgLeuAlaGluArgPheArg 495

73 CTTGGAGGTGTCTTTGCCCTTTCCTTCAGATGACCCACTAGGCGCTCC 26
   ::||||||| ||| |||||::: |||||||||
496 ......GlyArgPheArgAlaSer..ArgProProGlyArgArgSer 508

seq_name: pir2:S62706

seq_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) mito
C:Species: mitochondrion Platymonas subcordiformis
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 23-Jul-1999
C:Accession: S62706
R:Kessler, U.; Zetsche, K.
Plant Mol. Biol. 29, 1081-1086, 1995
A:Title: Physical map and gene organization of the mitochondrial genome from the unicell
A:Reference number: S62702; MUID:96145517
A:Accession: S62706
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-523 <KES>
A:Cross-references: EMBL:Z47795; NID:g633584; PIDN:CAA87753.1; PID:g633589
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
C:Genetics:
A:Gene: coI
A:Genome: mitochondrion
C:Function:
A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecule
ns from the mitochondrial matrix producing two molecules of water and lowering the conce
A:Pathway: oxidative phosphorylation; respiratory chain
A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-me
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium
ative phosphorylation; oxidoreductase; respiratory chain; transmembrane protein
F:1-13/Domain: mitochondrial matrix #status predicted <MM1>
F:13-460/Domain: cytochrome-c oxidase chain I homology <CO1>
F:14-42/Domain: transmembrane #status predicted <TM01>
F:43-54/Domain: intracristal #status predicted <ITC1>
F:55-90/Domain: transmembrane #status predicted <TM02>
F:91-98/Domain: mitochondrial matrix #status predicted <MM2>
F:99-121/Domain: transmembrane #status predicted <TM03>

F:122-144/Domain: intracristal #status predicted <ITC2>
F:145-174/Domain: transmembrane #status predicted <TM04>
F:175-186/Domain: mitochondrial matrix #status predicted <MM3>
F:187-216/Domain: transmembrane #status predicted <TM05>
F:217-231/Domain: intracristal #status predicted <ITC3>
F:232-265/Domain: transmembrane #status predicted <TM06>
F:266-272/Domain: mitochondrial matrix #status predicted <MM4>
F:273-289/Domain: transmembrane #status predicted <TM07>
F:290-301/Domain: intracristal #status predicted <ITC4>
F:302-330/Domain: transmembrane #status predicted <TM08>
F:331-338/Domain: mitochondrial matrix #status predicted <MM5>
F:339-359/Domain: transmembrane #status predicted <TM09>
F:360-373/Domain: intracristal #status predicted <ITC5>
F:374-403/Domain: transmembrane #status predicted <TM10>
F:404-409/Domain: mitochondrial matrix #status predicted <MM6>
F:410-436/Domain: transmembrane #status predicted <TM11>
F:437-449/Domain: intracristal #status predicted <ITC6>
F:450-481/Domain: transmembrane #status predicted <TM12>
F:482-522/Domain: mitochondrial matrix #status predicted <MM7>
F:65,381/Binding site: heme a iron (His) (axial ligands) #status predicted
F:244,293,294/Binding site: copper (His) #status predicted
F:244-248/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:248/Binding site: oxygen (Tyr) #status predicted
F:322/Binding site: myristate (Lys) (covalent) #status predicted
F:371/Binding site: magnesium (His) (shared with chain II) #status predicted
F:379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

alignment_scores:
Quality: 73.00   Length: 143
Ratio: 0.986   Gaps: 8
Percent Similarity: 51.748   Percent Identity: 25.874

alignment_block:
US-09-540-234-1 x S62706   ..

Align seg 1/1 to: S62706 from: 1 to: 523

92 GCTGCGACGGCATAATGGAAAACAGTCCCTGCCCTTTTGAGGTCGTCTC 141
   |||||||||::: ||| ||| || ::: ||| :::::||
126 AlaGlyThrGly...TrpThrValTyrProLeuSerSerIleAlaSe 141

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    ||:::   :::|||||| ::: :::::::|||
141 rHisSer......GlyAlaSerVal.AspLeuAlaIlePheSerLeuHis 155

192 AACGGCCAATGCGTCGATGTTATCGGAGCCATTCGTCGTGTCGTTTACTGT 241
    ||:::|||||||||:::|||||:::|||| ::: ||| |||:|:::
156 IleSerGlyAlaSerSerIleLeuGlyAlaLeuAsnPheIleThrThrIl 172

242 GTAGGCGTTG..............ATGACTTCCTTCGTCGTACTCCG 273
    : :::::||| ||| ||| |||||||
172 eLeuAsnMetArgGlyMetGlyMetThrHisArgLeuProLeuPheV 189

274 CTTGGAGCCCATTGAGTGCGTGTTACCGTGGCAGCACTCGAAGACTTCCTT 323
    :::||||| ||| :::|||||::: |||||||:::
189 alTrpSerValLeuIleThrAlaPheLeuLeuLeuSer......Leu 203

324 GCTGTACTCGCTGGCAGCATAGTCTGTTTGCTG............ 356
    ||||||||||:::||::::||::::|::: ||||||
204 ProValLeuAlaGlyAlaIleThrMetLeuLeuThrAspArgAsnPheAs 220

357 ..........CCTGGCTCGGGACACCTGTGTACTACAGTA 387
               ||||||| :: :::|||::: :::::::
220 nThrThrPhePheAspProAlaGlyGlyAspProIleLeuPheGlnH 237

388 CTTCGTTT...........CCCTTTGTAGTGGTACTACTTCTT... 419
    ||| ||| ::|||||:||||:|||
237 isLeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuPro 253

420 ..........AGCCATGTACTC 431
              |||||||:||
254 AlaPheGlyIleIleSerHisValVal 262

seq_name: pir2:S30291

seq_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - Prototheca wickerhamii mitochondrion
C;Species: mitochondrion Prototheca wickerhamii
C;Date: 19-Mar-1997 #sequence_revision 09-May-1997 #text_change 20-Aug-1999
C;Accession: S30291
R;Wolff, G.; Burger, G.; Lang, B.F.; Kueck, U.
Nucleic Acids Res. 21, 719-726, 1993
A;Title: Mitochondrial genes in the colourless alga Prototheca wickerhamii resemble plan
A;Reference number: S30291; MUID:93181269
A;Accession: S30291
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-514 <WOL>
A;Cross-references: EMBL:X68721; NID:g14210; PIDN:CAA48661.1; PID:g14211
C;Genetics:
A;Gene: coxI
A;Genome: mitochondrion
A;Introns: 129/2; 236/1; 239/3
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
transmembrane protein
F;10-456/Domain: cytochrome-c oxidase chain I homology <CO1>
F;62,377/Binding site: heme a iron (His) (axial ligands) #status predicted
F;240,290/Binding site: copper (His) #status predicted
F;240-244/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F;244/Binding site: oxygen (Tyr) #status predicted
F;367/Binding site: magnesium (His) (shared with chain II) #status predicted
F;375/Binding site: heme a3 iron (His) (axial ligand) #status predicted

alignment_scores:
        Quality: 72.50   Length: 145
        Ratio: 0.954   Gaps: 9
Percent Similarity: 52.414   Percent Identity: 26.897

alignment_block:
US-09-540-234-1 x S30291   ..

Align seg 1/1 to: S30291 from: 1 to: 514

 92 GCTGGCACGGCATAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTC 141
    ||||||||||| ||| ::: ||||:| :::|||||||||||||| :::|||
123 AlaGlyThrGly...TrpThrValTyrProLeuAlaSerIleAlaSe 138

142 AAACATGCCCCTCGAGACGTCGGTGAAGCCGTAGTTGTACAAGACTAAC 191
    |::: |||||| ::: :::::::::::::||||||||
138 rHisSer......GlyGlySerVal.AspLeuAlaIlePheSerLeuHis 152

192 AACGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGCTGTGTTTACTGT 241
    |||::   :::::::::::||||||||||: ::: |||||
153 LeuAlaGlyValSerIleLeuGlyAlaIleAsnPheIleCysThrVa 169

242 GTAGGCGTTG..............ATGACTTCCTTGCTGTACTCCG 273
    ||| ||| :::||||||:|||||||
169 lPheAsnMetAlaProGlyMetSerMetLeuAspLeuLeuPhe...V 185

274 CTTGGGAGCCCATTGAGTGCTGTTACCGTGCTGCAGCACTGCGAAGACTTCCTT 323
    ::||||: ::|||||||||:||||||||||||||::: |||
185 alTrpAlaValPheIleThrAlaTrpLeuLeuLeuLeu......CysLeu 199

324 GCTGTACTCGCTGCAGCAGCATAGTGCTGTTGCTG.............. 356
    |||||||||||::::|||||||::::::: |||||:
200 ProValLeuAlaGlyIleLeuThrMetLeuThrArgAspArgAsnPheAs 216

357 ...........CCTGGTCTGGGGACACCTGTGTTACTACAGTA 387
               |||||||:||||::::|||||||||||||:::
216 nThrSerPheAspProAlaGlyGlyAspProIleLeuTyrGlnH 233

388 CTTCGTTT..........CCCTTGTAGTGGTACTA........... 413
    ||| ||| ||:|||

seq_name: pir2:T20654

seq_documentation_block:
hypothetical protein F09C8.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T20654
R;McMurray, A.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z19306
A;Accession: T20654
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-779 <WIL>
A;Cross-references: EMBL:Z68132; PIDN:CAA92220.1; GSPDB:GN00028; CESP:F09C8.2
A;Experimental source: clone F09C8
C;Genetics:
A;Gene: CESP:F09C8.2
A;Map position: X
A;Introns: 22/2; 165/1; 266/3; 306/3; 376/1; 462/1; 729/2

alignment_scores:
    Quality: 72.50   Length: 128
      Ratio: 1.133   Gaps: 5
Percent Similarity: 50.000   Percent Identity: 21.094

alignment_block:
US-09-540-234-1/rev x T20654

Align seg 1/1 to: T20654 from: 1 to: 779

```
434 CACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGAAGTAC 385
    |||::::| ||| ::: |: ||: |||:::||:|
138 HisGlnGlnGlnGlnGlnHisGlnSerMetLysGlnGlnHi 154

384 TGTAGTACACAGGTGTCCCAGCAGCCAGGCAGCAATGCTGCCA 335
    :::::|| ||| :: |||||||||| ::|
154 sAlaMetHisLeu.....GlnGlnGlnProGlnThrAspTyrLeuSerS 169

334 GCGAGTACAGCAAGGAAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAA 285
    || ||| :::::::::::
169 er..............GluValCysTyrProSerThrSerSer 178

284 ATGGCTCCAAGCGAGCGAGTACAGCAAGGAAGTCATCAAGCGCTACACAGTAA 235
    ||||||||| :::::::: ::|||
179 TyrGluProSerGlyTyrArgGluSerGlyTyrCysProGlnAsnSerTh 195

234 A...........CAGCACACGAATGGCTCCGATA 212
    ||||| ||| :||||
195 rIleLeuProAspAspProGlnAsnTyrGlnHisGlyAsnAspGlyAspA 212

211 ACATCGACGCCATTGGC..................... 195
    :: :::::|||
212 spGluGluSerPheGlyGluProGluArgTyrGluProMetSerProThr 228

194 ......GTTGTTAGTCTTGTACAAACTACGGCCTCACCGACGTCTCGAGG 151
       ||||||||| ::||||
229 PheSerIleHisSerPheAlaProThrThrSerThrAsnThrAlaGluAl 245

150 G........GCATGTTTGAGACGACCCTCA 129
       ::||||| ::|
245 aGluAspLeuGluValCysIlePhePheProAla 256
```

seq_name: pir2:S28066

seq_documentation_block:
sexual differentiation protein Esc1p - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 31-Jan-2000
C;Accession: S28066; T38926
R;Benton, B.K.; Reid, M.S.; Okayama, H.

EMBO J. 12, 135-143, 1993
A;Title: A Schizosaccharomyces pombe gene that promotes sexual differentiation encode
A;Reference number: S28066; MUID:93154314
A;Accession: S28066
A;Molecule type: mRNA
A;Residues: 1-413 <BEN>
A;Cross-references: EMBL:X69389; NID:g4949; PIDN:CAA49186.1; PID:g4950
R;Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajendream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A;Reference number: Z21817
A;Accession: T38926
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-413 <PEA>
A;Cross-references: EMBL:Z69728; PIDN:CAA93587.1; GSPDB:GN00066; SPDB:SPAC56F8.16
A;Experimental source: strain 972h-; cosmid c56F8
C;Genetics:
A;Gene: escl; SPAC56f8.16
A;Map position: 1
C;Keywords: DNA binding; nucleus

alignment_scores:
    Quality: 72.00   Length: 97
      Ratio: 1.180   Gaps: 1
Percent Similarity: 62.887   Percent Identity: 24.742

alignment_block:
US-09-540-234-1/rev x S28066

Align seg 1/1 to: S28066 from: 1 to: 413

```
398 GGGAAACGAAGTACTGTAGTACACAGGTGTCCCAGCCAGGCAGCAAAC 349
    |||::: |||::: |||:: |||:::|| :::::||:
115 GlySerSerSerTyrValTyrSerValProProThrAsnSerThrTh 131

348 AGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTCGAGTGCTGCACGG 299
    :::: |||:|||||||||| ::
131 rSerGlnAlaSerAlaLysHisSerAlaValProHisArgSerSerGlnP 148

298 TAAACAGCACTCAAATGGCTCCAAGCGAGTACAGCAAGGAAGTCATCAA 249
    ::|||||: :::|::::|||||| :::::::::::
148 heGlnSerThrThrLeuThrProSerThrThrAspSerSerSerThrAsp 164

248 GCGCTACACAGTAAACAGCACGAATGGCTCCGATAACATCGACGCCAT 199
    :: :||| :::::::::::||||
165 ValSerSerAspSerValSerThrSerAlaSerSerSerAsnAlaSer 181

198 TGGCGTTGTTAGTCTTGTACAAACTACGGCCTCACCGACGTCTCGAGGGG 149
    :::::|||||::: :: :||||
181 rAsnThrValSerVal.......ThrSerProAlaSerSerAlaThrP 196

148 CATGTTTGAGACGACCCTCAAAAGGGCAGGGACTGTTTTCC 108
    :::||||||:: ||||:||
196 roLeuProAsnGlnProSerGlnGlnGlnPheLeuValSer 209
```

seq_name: pir2:T11317

seq_documentation_block:
cytochrome-c oxidase (EC 1.9.3.1) chain I - Pedinomonas minor mitochondrion
C;Species: mitochondrion Pedinomonas minor
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jan-2000
C;Accession: T11317
R;Turmel, M.; Lemieux, C.; Burger, G.; Lang, B.F.; Otis, C.; Plante, I.; Gray, M.W.
submitted to the EMBL Data Library, December 1998
A;Description: The complete mitochondrial DNA sequences of Nephroselmis olivacea and
A;Reference number: Z17261
A;Accession: T11317
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-523 <TUR>
A;Cross-references: EMBL:AF116775; NID:g4378766; PID:g4378771; PIDN:AAD19669.1
C;Genetics:

A;Gene: coxI
A;Genome: mitochondrion
A;Genetic code: SGC3
C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C;Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated comple
F;12-459/Domain: cytochrome-c oxidase chain I homology <CO1>
F;64,380/Binding site: heme a iron (His) (axial ligands) #status predicted
F;243,292,293/Binding site: copper (His) #status predicted
F;243-247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F;247/Binding site: oxygen (Tyr) #status predicted
F;370/Binding site: magnesium (His) (shared with chain II) #status predicted
F;378/Binding site: heme a3 iron (His) (axial ligand) #status predicted

alignment_scores:
            Quality: 72.00   Length: 161
            Ratio: 0.986   Gaps: 7
Percent Similarity: 45.342   Percent Identity: 24.845

alignment_block:
US-09-540-234-1 x T11317   .

Align seg 1/1 to: T11317 from: 1 to: 523

107 TGGAAACAGTCCTGCCCTTTTGAGGGTGTCTCAAACATGCCCCTCGA 156
    ||| ||::|||||||| :::||||:::::||| :::||
106 TrpLeuLeuProProSerLeuLeuLeuValGlySerMetTyrIleGl 122

157 G..................................... 157
122 uIleGlyAlaGlyThrGlyTrpThrValTyrProProLeuSerLeuIleG 139

158 ...........ACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAA 193
            :::||||||::| :|| ||||| ::||||||||
139 luPheHisSerAlaSerVal.AspMetAlaIlePheSerLeuHisVa 155

194 CGCCAATGGCGTCGATGTTATCGGAGCCATTCGTGTGTCGTTTACTCGT 243
    ::::|||||||||||||||||| :::|| |||:::
155 lSerGlyLeuSerLeuGlyAlaIleAsnPheIleValThrIleP 172

244 AGCGCTTGATGACT............TCCTTGCTGACTCCGCT 275
    ::: ||||||||||||| :::||||
172 heCysMetLysThrArgGlyLeuSerTrpArgAlaLeuProLeuPheVal 188

276 TGGAGCCATTGAGTCGTGTTACCGTGCAGCACTCGAAGACTTCCTTGC 325
    |||||| || :::::|| || |||
189 TrpSerValLeuThrAlaIleLeuLeuLeuLeuThr......LeuPr 203

326 TGTACTCGCTGGCAGCATAGTCTGTTTGCTGCCT................. 359
    |||||||||||||::: ||||||
203 oValLeuAlaGlySerLeuThrMetLeuLeuThrAspArgHisPheAsnT 220

360 ..........GGTCTGGGGACACCTGTGTACTACAGTACT 389
                |||||| ||| :::||::::
220 hrSerPheTyrSerValTyrGlyValGlyAspProValLeuTyrGlnHis 236

390 TCGTTT..........CCCTTTGTAGTGGTACTACTTCTT...... 419
    || ||| :::||:::|||
237 LeuPheTrpPhePheGlyHisProGluValTyrIleLeuIleLeuProAl 253

420 ..........AGCCATGTACTCGTGTCT 437
              ||||||||||||:::|||
253 aPheGlyIleIleSerGlnValIleIleSer 263

seq_name: pir2:T13646

seq_documentation_block:
hypothetical protein 95B7.9 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13646
R;Ferraz, C.; Vidal, S.; Brun, C.; Bucheton, A.; Demaille, J.G.
submitted to the EMBL Data Library, April 1999

A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17694
A;Accession: T13646
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-735 <FER>
A;Cross-references: EMBL:AL021728; PIDN:CAA16821.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0000376
A;Introns: 39/3; 93/2; 294/2; 387/2
A;Note: EG:95B7.9
C;Superfamily: fruit fly hypothetical protein EG_95B7.9

alignment_scores:
            Quality: 72.00   Length: 89
            Ratio: 1.241   Gaps: 5
Percent Similarity: 65.169   Percent Identity: 25.843

alignment_block:
US-09-540-234-1/rev x T13646   .

Align seg 1/1 to: T13646 from: 1 to: 735

408 CCACTACAAAGGGAAACGAGTACTGTAGTACACAGGTGTCCCAGACCA 359
    |||||||:::::||||| :::||||||| :::||||||::::
372 ProLeuGluHisAspMetLys.LeuLeuIleGlnArgCysProAsnIleL 388

358 GGCAG...CAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAAGTCTTC 312
    :::||| || |||| :::
388 ysAsnTyrGluValAspPheValSerGlyTyrGlnLysAlaLeuAla 404

311 GAGTGCTGCACG...GTAAACAGCACTCAAATGGCTCCAAGCGGAGTACA 265
    :::|||| ||| |:::||||||| ||||||||||
405 SerCysProAsnGlyIleMetAlaSerAspMetAlaSerAsnGlyValGl 421

264 GCAAGGAAGTCATCAAGCCGCTACACAGTAAACAGCACACGAATGGCTCCG 215
    | :::||| :::|||||| :||||||||||||||
421 nValThrGlyAsn.......ValGlnAspGluGluPheLysAsnGlyAlapS 436

214 ATAACATCGACGCCATTGGCGTGTTGTAGTCTTGTACAAACTACGGCTTCA 165
    ::: |||||||| ||| |||
436 er.....................SerArgThrSerProGlnProSer 444

164 CCGACGGTCTCGAGGG 150
    |||||||||:::||
445 ProThrAlaAsnGly 449

seq_name: pir2:T25005

seq_documentation_block:
hypothetical protein T19C9.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T25005
R;Matthews, L.
submitted to the EMBL Data Library, March 1997
A;Reference number: Z19967
A;Accession: T25005
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1221 <WIL>
A;Cross-references: EMBL:Z92972; PIDN:CAB07490.1; GSPDB:GN00023; CESP:T19C9.7
A;Experimental source: clone T19C9
C;Genetics:
A;Gene: CESP:T19C9.7
A;Map position: 5

alignment_scores:
            Quality: 72.00   Length: 66
            Ratio: 1.600   Gaps: 2
Percent Similarity: 68.182   Percent Identity: 28.788

alignment_block:
US-09-540-234-1/rev x T25005 ..

Align seg 1/1 to: T25005 from: 1 to: 1221

290 ACTCAAATGGCTCCAAGCGGAGTACAGCAGGAAGTCATCAAGCGCTACA 241
    ::|||||: ||| |||||:::|||:   :: |||||
 81 SerGlnIleMetIleGlyLyLysLeuGlnSerGluLeuSerLysLeuHi 97

240 C.........AGTAAACAGCACACGATGGCTCCGATAACATCG 206
    ||||||::: :: :|||:|||||||||
 97 sLeuGluGlyGluLysSerLysGluLysSerGlyValSerGlyThrAspAsnIleA 114

205 ACGCCATTGGCGTTGTTAGTCTTGTTACAAACTACGGCTTCACCGACGTCT 156
    ||::: ||| :: ||| ::: |||||||||
114 spCysAlaSerProLeuSerLeu...ArgLysValArgSerProThrIle 129

155 CGAGGGGCATGTTTGAGACGACCAACCCTCAAAAGGGCAGGAGGACTGTTTTCC 108
    :::::||||| ||||||::: |||:::|||:|||||
130 AsnSerValPheLeuLysLysProAlaAlaAsnArgAlaLeuTyrThr 145

seq_name: pir2:T36052

seq_documentation_block:
probable cytochrome-c oxidase chain II - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T36052
R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z21595
A;Accession: T36052
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-334 <SAU>
A;Cross-references: EMBL:AL034355; PIDN:CAA22218.1; GSPDB:GN00070; SCOEDB:SCD78.13
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: cydB; SCOEDB:SCD78.13
C;Superfamily: cytochrome d ubiquinol oxidase

alignment_scores:
        Quality: 71.50   Length: 161
        Ratio: 1.135   Gaps: 4
Percent Similarity: 39.130   Percent Identity: 19.876

alignment_block:
US-09-540-234-1 x T36052 ..

Align seg 1/1 to: T36052 from: 1 to: 334

  6 CACCGGGCCCTTATGTATGCAGGAGGCGCCTAGCTGGTGGTCATCTGAAGGAAAG 55
    |||||| :: ||||||::::|||::||:||
175 HisGlyThrValPheThrAlaLeuLysThrValGlyGluThrArgGluAr 191

 56 GACAAAGACACCTCCAAGCCGGACACGGCTACTG................ 89
    |:::::||| |||::
191 gAlaArgThrLeuAlaAlaArgValGlyLeuValThrAlaValLeuValAlaLeuAlaL 208

 89 89
208 euAlaPheLeuLeuTrpThrGlnAlaAspSerGlyAspAlaLysSerLeu 224

 90 ....TAGCTGGCACG 100
    :: |||::
225 ValAlaLeuValValAlaValAlaAlaLeuValAlaAlaLeuMetAlaAs 241

101 GCATAATGGAAAACAGTCCCTGCCCTTTTGAGGGTCGTCTCAAACATGCC 150
    :::| ||::::::: |||
241 nGlnAlaGlyArgGluGlyTrpSerPhe................... 250

151 CCTCGAGACGTCGGTGAAGCCGTAGTTTGTACAAGACTAACAACGCCAAT 200
                                           ::
251 ....AlaLeuSer 253

201 GGCGTCGATGTTATCGGAGCCATTCGTGTCGTGTTTACTGTG......TA 244
    ||||| :::|||||||:::||||:   :::
254 GlyValThrIleValAlaAlaValAlaMetLeuPheLeuThrLeuPhePr 270

245 GCGCTTGATGACTTCCTTGCTGTACTCCGCTGGAGC.............. 281
    ::::||||||||| ||| ::: |||||
270 oAsnValMetThrSerThrLeuAsnAlaAspTrpSerLeuThrValThrA 287

282 ..CATTTGAGTGCTGTTTACCGTGCAGCACTCGAAGACTTCCTTGCTGTA 329
    ||::|||||| ::||||||||
287 snAlaSerSerSerAlaTyrThrLeuLysIleMetThrTrpLeuAlaVal 303

330 CTCGCTGGCAGCATAGTCTGTTTGCCCTGT 362
    ::||| :|| ||| |||
304 IleAlaThrProValValLeuTyrGlnGly 314

seq_name: pir2:A44265

seq_documentation_block:
trithorax homolog HTX, version 2 - human
N;Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)
C;Species: Homo sapiens (man)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 01-Dec-2000
C;Accession: A44265; A44264; I58112; I37165; I38485
R;Tkachuk, D.C.; Kohler, S.; Cleary, M.L.
Cell 71, 691-700, 1992
A;Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal transl
A;Reference number: A44265; MUID:93046667
A;Accession: A44265
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-3968 <TKA>
A;Cross-references: GB:L04284; NID:g184393; PIDN:AAA58669.1; PID:g184394
A;Note: sequence extracted from NCBI backbone (NCBIP:117729)
R;Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Canaani, O.; Cimino, G.; Croce, C.M.;
Cell 71, 701-708, 1992
A;Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-
A;Reference number: A44264; MUID:93046668
A;Accession: A44264
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 63-316,'GLLINSELEK',327,'Q',329,'VR',332,'DKEGTPP',340,'T',342,'EDKTVRQS
546,'LQIESTSP',2555-3554,'N',3556-3594,'V',3596-3899,'A',3901-3968 <GU1>
A;Cross-references: GB:L04731; NID:g339921
A;Note: sequence extracted from NCBI backbone (NCBIP:117779)
R;Djaball, M.; Selleri, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.
Nature Genet. 2, 113-118, 1992
A;Title: A trithorax-like gene is interrupted by chromosome 11q23 translocations in a
A;Reference number: I58112; MUID:93265134
A;Accession: I58112
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1317-1700,'DD',1703-1936,'H',1938-2180,'S',2182-2328 <DJA>
A;Cross-references: GB:L01986; NID:g307522; PIDN:AAA92511.1; PID:g553800
R;Marschalek, R.; Greil, J.; Lochner, K.; Nilson, I.; Siegler, G.; Zweckbronner, I.;
Br. J. Haematol. 90, 308-320, 1995
A;Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in t
A;Reference number: I37165; MUID:95315013
A;Accession: I37165
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1212-1603,'GTE','MAR>
A;Cross-references: EMBL:X83604; NID:g897757; PIDN:CAA58584.1; PID:g899268
A;Note: submitted to the EMBL/GenBank/DDBJ databases by R. Marschalek, 20 December 19
R;Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Canaani, O.; Saito, H
Cancer Res. 54, 2327-2330, 1994
A;Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involve
A;Reference number: I38485
A;Accession: I38485

A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1251-1486,'G',1488-1538 <RES>
A;Cross-references: EMBL:U04737; NID:g451554; PID:g451555
C;Genetics:
A;Gene: GDB:MLL; HTX; ALL-1; HRX
A;Cross-references: GDB:128819; OMIM:159555
A;Map position: 11q23-11q23
A;Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
A;Note: the list of introns is incomplete
C;Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
F;1434-1456/Region: zinc finger CCHC motif
F;1479-1506/Region: zinc finger CCHC motif
F;1527-1556/Region: zinc finger CCCC motif
F;1569-1596/Region: zinc finger CCHC motif
F;1873-1900/Region: zinc finger CCHC motif
F;1933-1955/Region: zinc finger CCHC motif

alignment_scores:
         Quality: 71.50   Length: 118
         Ratio: 1.192   Gaps: 6
Percent Similarity: 50.847   Percent Identity: 29.661

alignment_block:
US-09-540-234-1/rev x A44265

Align seg 1/1 to: A44265 from: 1 to: 3968

367 CCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGAGTACAGCAAGGAA 318
    |||::: |||||||||||||| ::|||| |||:||
3406 ProGlnLeuGlyThrSerGlnThrProSerThrAlaAlaIleThrAlaAl 3422

317 GTCTTCCAGTGCTGCACGTAAACAGCACTCAA................. 285
    |||||| |||:::::: ||||||||
3422 aSerSer..IleCysValLeuProSerThrGlyThrThrGlyIleThrAl 3438

284 .ATGGCTCCAAGCGGA.......................GTAC 266
     |||||||||| |||
3438 aAlaSerProSerGlyGluAlaAspGluHisTyrGlnLeuGlnHisValA 3455

265 AGCAA....GGAAGTCATCAAGCGCTACACAGTAAACAGCAC...... 228
    ::||| :||| |||||::||::::|||||||::|||||
3455 snGlnLeuAlaSerLysThrGlyIleHisSerSerGlnArgAspLeu 3471

227 ...................ACGAATGGCTCCGATAACATCGACGC 202
                       :||| |||||||::||:::|||||
3472 AspSerAlaSerGlyProGlnValSerAsnPheThrGlnThrValAspAl 3488

201 CATTGGCGTTGTTAGTCTTGTACAA..................ACTA 173
    | ||:|::||:||| ||:
3488 aProAsnSerMetGlyLeuGluGlnAsnLysAlaLeuSerSerAlaValG 3505

172 CGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACGACCCTCAAAAGGG 123
    ||||||||||| ::||| |||:::||
3505 lnAlaSerProThrSerProGlyGlySerProSerSerSerGly 3521

122 CAG 120
    |||
3522 Gln 3522

seq_name: pir2:T07068

seq_documentation_block:
beta-fructofuranosidase (EC 3.2.1.26) LIN7 - tomato (fragment)
N;Alternate names: invertase
C;Species: Lycopersicon esculentum (tomato)
C;Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
C;Accession: T07068
R;Godt, D.E.; Roitsch, P.
Plant Physiol. 115, 273-282, 1997
A;Title: Regulation and tissue-specific distribution of mRNAs for three extracellular in
A;Reference number: Z15897; MUID:97451780

A;Accession: T07068
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-244 <GOD>
A;Cross-references: EMBL:X91391; NID:g1771153; PIDN:CAA62736.1; PID:g1771154
A;Experimental source: cultivar Moneymaker; leaf
C;Genetics:
A;Gene: LIN7
C;Superfamily: beta-fructofuranosidase
C;Keywords: cell wall; glycosidase; hydrolase

alignment_scores:
         Quality: 71.00   Length: 96
         Ratio: 1.340   Gaps: 3
Percent Similarity: 55.208   Percent Identity: 26.042

alignment_block:
US-09-540-234-1/rev x T07068

Align seg 1/1 to: T07068 from: 1 to: 244

427 ACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGAA...GTACTGTA 381
    |||:||| ||
48 ThrTrpSerGlySerAlaThrIleLeuProGlyAsnLysProValIleLe 64

380 GTACACAGGTGTCCCCAGACCAGGCAGCAAACAGACTATGCTGCCAGCGA 331
    |||||||||| ||::::: |||:::::: ||
64 uTyrThrGlyIleIleAspAlaAsnGlnThrGlnValGlnAsnTyrAlaI 81

330 GTACAGCAAGGAAGTCTTCGAGTGCTGCACGTAAACAGCACTCAAATGG 281
    |||:|||| |||::: ||| || ::|||
81 leProAlaAsnLeuSerAspProTyrLeuArg..........GluTrp 93

280 CTCCAAGCGGAGTACAGCAAGGAAGTCATCAAGCGCTACACAGTAAACAG 231
    ||||||| ||:||| ::|||:::::|||:::::|||
94 IleLysProAspAsnProLeuIleIleAlaAspGluSerIleAsnLy 110

230 CACACGAATGGCTCCGATAACATCGACGCCATTGGCG............. 194
    :||::::: |||:|| :::::
110 sThrLysPheArgAspProThrAlaTrpMetGlyLysAspGlyHisT 127

193 ..TTGTTAGTCTTGTACAAACTACGGCTTCACCGACGT 158
        :||:||| :::|||||| |||||| ||
127 rpArgIleValMetGlySerLeuArgLysHisSerArg 139

seq_name: pir2:A83867

seq_documentation_block:
homoserine dehydrogenase BH1737 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C;Accession: A83867
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A;Reference number: A83650; MUID:20263314
A;Accession: A83867
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-342 <STO>
A;Cross-references: GB:AP001513; GB:BA000004; NID:g10174345; PIDN:BAB05456.1; GSPDB:G
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH1737

alignment_scores:
         Quality: 72.00   Length: 89
         Ratio: 1.220   Gaps: 2
Percent Similarity: 66.292   Percent Identity: 30.337

alignment_block:

US-09-540-234-1/rev x A83867

Align seg 1/1   to: A83867   from: 1   to: 342

```
323 AAGGAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCTCCAAG 274
    ::|||: ||| ::: :||||::::::::||||:|||||:::
134 GlnGluLeuLysGluAlaAsnGlyValGlnLeuLeuTyrGl 150

273 CGGAGTACAGCAAGGAAGTCATCAAGCGCTACACAGTAAACAGCACACGA 224
    :|||:|| ::::::::: |||||||   ::|||||||
150 uGlyThrValMetSerGlyThrProAlaLeuArgLeuProGluHisThrL 167

223 ATGGCTCCGATAACATCGACGCCATT...GGCGTTGTTAGTCT-TGTACA 178
    ::::::||||||||||||::::::||          |||
167 euIleGlyAsnAsnIleGlnLysGlyValLysGlyIleLeuAsnGlyThrThr 183

177 AACTACGGCTTCACCGAGCTCTCGAGGGC...ATGTTTGAGACGACCCT 131
    |||||| ::|||||::||::|||||| |||| ||||| ::|||
184 AsnTyrIleLeuThrGluMetGluArgGlyLysThrPheGluGlnAlaLe 200

130 CAAAAGGGCCAGGGACTGTTTCCATTATCCGTGCCAGCTACAGTAGCCG 81
    |::: ||| |||::: ::: :|| |||    |
200 uGlnAspAlaGlnThrLeuGlyTyrAlaGluAlaAspProThrSerAspV 217

80 TGTCCGGCTTGGAGGTG 64
   ||::|||   ::|||
217 alGluGlyTyrArgAspVal 222
```

seq_name: pir2:D96810

seq_documentation_block:
hypothetical protein T1I11.6 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96810
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: D96810
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-530 <STO>
A;Cross-references: GB:AE005173; NID:g6587860; PIDN:AAF18546.1; GSPDB:GN00141
C;Genetics:
A;Gene: T1I11.6
A;Map position: 1

alignment_scores:
              Quality: 71.00   Length: 64
              Ratio: 1.651   Gaps: 2
Percent Similarity: 67.188   Percent Identity: 32.812

alignment_block:
US-09-540-234-1/rev x D96810   ..

Align seg 1/1   to: D96810   from: 1   to: 530

```
439 ACAGACACGGAGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGA 390
    ||||||||||||| ||| :::||| ||| :::|||
65 ThrAspThrGluTyrSerSerGlnProLysSerAspSerGlnLysProGlnGl 81

389 AGTACTGTAGTACACAGGTGTCCCAGACCAGCCAGCAAACAGACTATGC 340
    ::::|||     :::|::|||  |||  |||
81 n.................LysProLysProAspGluAsnHisProArgL 92
```

US-09-540-234-1/rev x A83867

Align seg 1/1   to: A83867   from: 1   to: 342

```
339 TGCCACGGAGTACAGCAAGGAAGTCTTCGAGTGCTGCACGGTAAACAGCA 290
    ::::::::|||||||||||||||||:|||:::::|||||| ||||:
92 ysSerSerGluSerAlaArgLysSerSerAspSerAlaArg.LysSerIl 108

289 CTCAAATGGCTCCAAGCGGAGTACAGCAAGGAAGTCATCA 250
    |||:::||||:||||::|||: :::::::::
108 eSerSerGlySerSerArgThrGluSerLysArgPheSer 121
```

seq_name: pir2:T45785

seq_documentation_block:
hypothetical protein F26O13.180 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T45785
R;Delseny, M.; Berger, C.; Cooke, R.; Grellet, F.; Laudie, M.; Mewes, H.W.; Lemcke, K
submitted to the Protein Sequence Database, December 1999
A;Reference number: Z23013
A;Accession: T45785
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-438 <DEL>
A;Cross-references: EMBL:AL133452
A;Experimental source: cultivar Columbia; BAC clone F26O13
C;Genetics:
A;Map position: 3
A;Introns: 35/1; 67/2; 104/3; 326/3
A;Note: F26O13.180

alignment_scores:
              Quality: 70.50   Length: 166
              Ratio: 0.860   Gaps: 9
Percent Similarity: 49.398   Percent Identity: 27.108

alignment_block:
US-09-540-234-1/rev x T45785   ..

Align seg 1/1   to: T45785   from: 1   to: 438

```
436 GACACGGAGTACAGCGGCTAAGAAGTAGTACCACTACAAAGGGAAACGAAGT 387
    ::::|||::::||::: |||:::||| ||||||||||||||:
88 AsnThrThrSerLysLeuGlnSerIleThrThrThrArg............... 100

386 ACTGTAGTACACAGGTGTCCCC.....................AGACCAGGCAGCA 352
    ||||||                     |||:::|||
101 ............ValProLeuArgSerGluSerAspProSerSerA 112

351 AACAGACTATCGTGCCAGCAGTACAGCAAG................ 320
    ::  ||| ||||||||||||||||||
112 rgProThrArg...SerGlySerThrIleArgProSerAsnIleProThr 127

319 ...AAGTCTTCGAGTGCTGCACGGTAAACAGCACCACCAAATGGCTCCAAGC 273
    :::||||||||||  ::: ||| :::
128 IleArgSerSerSerValProLysLysThrThrThrThrGlnIleGlnAl 144

272 GGAGTACAGC..........AAGGAAGTCATCAAGCGCTACACAGTAA 235
    |::: |||          ::::|||||||| :::
144 aSerAlaSerValSerProLysArgThrValSerArgSerLeuThrP 161

234 ACAGACACGAATGGCTCCGATA...ACATCGACGCCATTGGCGTTGTTA 188
    ||:::||| ::::|||   ::: :||
161 roSerSerArgLysThrProSerProThrSerProSerArgIleSer 177

187 GTCTTGTACAAACTACGGCTTCACCGAGCGTCTCGAGGGCATGTTTGAGA 138
    ::::               |||
178 ThrThrThrSerThrProSerPheLysThrAlaGly............. 190

137 CGACCCTCAAAAGGGCAGGGACTGTTTCCATTATGCCGT............ 98
    |||  :::::::                |||
191 .AspAlaGlnArgSerArgSerLeuThrProArgAlaLysProGlnIleA 207
```

```
 97 ..GCCAGCTACAGTAGCCGTGTCCGGCTTGGAGGTGTCTTTGTCCTTCC  50
      |||::: |||||||||:|::::::       :::  ||   |||   |||
207 laAlaAsnAlaSerSerArgThrAsnValArgSerSerValThrSer 223

 49 TTCAGATGACCCATCAGGCGCTCCTGCATACATAAGGGCCCGTGGACC   2
    ||| ||| |||      :::      ::::::      |||   |||
224 Arg......ProAspArgSerThrThrValSerAlaThrProArgThr 237

seq_name: pir2:T48357

seq_documentation_block:
hypothetical protein F12E4.90 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T48357
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-684 <BEV>
A;Cross-references: EMBL:AL162751
A;Experimental source: cultivar Columbia; BAC clone F12E4
C;Genetics:
A;Map position: 5
A;Note: F12E4.90

alignment_scores:
         Quality:       70.50           Length:         101
         Ratio:          1.306          Gaps:             6
Percent Similarity:    53.465       Percent Identity:   23.762

alignment_block:
US-09-540-234-1 x T48357

Align seg 1/1  to: T48357  from: 1 to: 684

172 GTAGTTGTACAAGACTAACAACGCCAATGGCGTCG.............    207
    ||| |||::  ||| :::::||  ::|||  |||
521 ValGlnCysAlaValValSerGluProLeuValHisGluSerHisMetH1 537

208 .........ATGTTATCGGAGCCATTCGTGTGC.....TGTTTACTGT    241
             |||||||||:|::::||:||       ::|| :|
537 sProLeuPheLeuThrSerGluProTyrGluAlaArgLeuCysSerValC 554

242 GTAGCCGTTGATGACTTCCTTGCTACTCCGCTTGGAGCCATTGGAGTG    291
    |||:::::||  ||||| ||:::|||  |||:::||
554 ysLysLys.GluSerTrpLeuLeuSerThrLysGluThrPheAsnCy 570

292 C...........................TGTTTACCGTGCAGCACTCGAAGAC    317
                               ||:::: |||:::|| ||:::||
570 sIleGluGluCysAspPheAlaLeuCysPheGlyCysAlaThr......L 585

318 TTCCTTGCTACTCCGCTGGCAGCATAGT.........CTGTTTGCTGCC    358
    |||||  |||:::|||   ||||         ::  ||
585 euProGlnLysLeuArgTyrLysHisAspLysHisMetLeuThrLeuSer 601

359 TGGTCTGGGGACACCTGTGTACTACAGTACTTCGTTTCCCTTGTAGTGG    408
    ::::::: |||   ||:|::| ||:::||||:::|::|
602 TyrGlyAsnLysThrSerThrMetAlaTyrTrpCysGluValCysGluGl 618

409 T 409
618 y 618

seq_name: pir2:T34433

seq_documentation_block:
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
```

```
C;Accession: T34433
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34433
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1032 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9.1a
C;Genetics:
A;Experimental source: strain Bristol N2; clone K06A9
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2;

alignment_scores:
         Quality:       70.50           Length:         123
         Ratio:          0.966          Gaps:             2
Percent Similarity:    59.350       Percent Identity:   26.016

alignment_block:
US-09-540-234-1/rev x T34433

Align seg 1/1  to: T34433  from: 1 to: 1032

439 ACAGACACGAGTACATGGCTAAGAAGTAGTACCACTACAAAGGGAAACGA  390
    ||| |||||    ||||||||     ||||||| :::::||| :::
698 ThrLeuThrSerGluProSerSerThrGlnSerSerGlyAlaGl 714

389 AGTACTGTAGTACAAGGTGTCCCCAGACCAGGCAGCAAACAGACTATGC  340
    ||| :::|||||||:|::::::   ::|| |||:::||
714 nSerThrLeuThrThrProSerProAsnProSerGlnSerThrSerSerL 731

339 TGCCAGCGAGTACGACCAGGAAGTCTTCGAGTGCTGC.ACGGTAAACAGC  291
    ||  :::|||||:::   |||:::|| |||:::::::::
731 euGluSerSerThrSerGlyAlaThrThrSerGlySerAlaGlyThr 747

290 ACTCAAATGGCTCCAAGC................GGAGTACAGCAAGGAAG  256
    ||:::|||||      :::||||||   :: ||||||||
748 ThrMetThrSerProSerGlnSerSerSerValGlySerSerGlnGlySe 764

255 T.........................CATCAAGGCGCTACACAGTA  236
    :|||||| ::::::      ||
764 rThrSerProAlaAlaSerThrThrSerGlyGluMetThrSerGlnGlys 781

235 AACAGCACACGAATGGCTCCGATAACATCGACGCGCCATTGGCGTGTTAGT  186
    :: :::||| ||:::|| |||:::||:::
781 erThrGlnThrProGlySerValSerThrSerAlaAlaIleLeuThr 797

185 CTTGTACAAACTACGGCTTCACCGAGCTCGAGGGGCATGTTGAGACG  136
    :::|| |||:::|| ||:|| ||:::  |:: ||
798 SerThrGlnGlnSerValSerThrAsnSerProGlySerThrValThrAr 814

135 ACCCTCAAAAGGGCAGGGA 117
    |||||    ::::||
814 gProSerThrValSerGly 820

seq_name: pir2:S52957

seq_documentation_block:
bimD protein - Emericella nidulans
C;Species: Emericella nidulans, Aspergillus nidulans
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 07-May-1999
C;Accession: S52957
R;Denison, S.H.; Kaefer, E.; May, G.S.
Genetics 134, 1085-1096, 1993
A;Title: Mutation in the bimD gene of Aspergillus nidulans confers a conditional mito
A;Reference number: S52957; MUID:93387663
A;Accession: S52957
A;Status: preliminary
A;Molecule type: mRNA
```

A:Residues: 1-1506 <DEN>
A:Cross-references: EMBL:L03200; NID:g168024; PID:g168025

alignment_scores:
        Quality: 70.50   Length: 119
        Ratio: 1.119   Gaps: 4
Percent Similarity: 52.941   Percent Identity: 23.529

alignment_block:
US-09-540-234-1/rev x S52957

Align seg 1/1 to: S52957 from: 1 to: 1506

 378 ACACAGGTGTCCCGACCACCAGGCAGCCAAACAGACTATGCTGCCAGCGAGT 329
     || ::||||||||:::|||||| ::||||
1283 ThrThrPheLeuProGluAspAlaGluAspArgLeu.....AspArgIl 1297

 328 ACAGCCAAGGAAGTCTTCGAGTGCTGCACGGTAAACAGCACTCAAATGGCT 279
     :::||||||:::||||: ||:|||||| :::||||||||||||
1297 eValArgAlaThrMetArgThrLysAsnGlySerGlnAlaGlyAlaL 1314

 278 CCAAGCGGAGTACA........   265
     ::|||||:::|||
1314 ysLysArgLysThrAspThrGlnGluProSerArgAspAlaAsnAla 1330

 264 ......GCAAGGAAGTCATCA..   250
           ||||||||||:::|||
1331 AlaLysLysAlaArgLysAsnSerGluSerSerSerArgLysSerGl 1347

 249 ......................AGCGCTACACAGTAAA   234
                           ::|||||||||:::| |
1347 ySerSerAlaValGlyPheLysMetProLysArgLysMetProLysLysT 1364

 233 CAGCACCACGAATGGCTCCGATAACATCGACGCCATTGCCGTTGTTAGTCT 184
     ||:: :::||| ||::: ::||| || |||:::||||||||
1364 hrGlyAspAspTrp.SerSerAspGlyGluAlaAlaGlyAsnValAlaSe 1380

 183 TGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACGAC 134
     ::::||||||||||:::||||||||||:::|||||||
1380 rSerSerAlaThrArgArgArgSerAsnArgGlySerAlaSerArgArgI 1397

 133 CCTCA 129
     |||
1397 leSer 1398

seq_name: pir2:T34434

seq_documentation_block:
hypothetical protein K06A9.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T34434
R;Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A;Description: The sequence of C. elegans cosmid K06A9.
A;Reference number: Z21525
A;Accession: T34434
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2232 <GEI>
A;Cross-references: EMBL:U80846; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A;Experimental source: strain Bristol N2; clone K06A9
C;Genetics:
A;Gene: CESP:K06A9.1a
A;Map position: X
A;Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/

alignment_scores:
        Quality: 70.50   Length: 123
        Ratio: 0.966   Gaps: 2
Percent Similarity: 59.350   Percent Identity: 26.016

alignment_block:
US-09-540-234-1/rev x T34434

Align seg 1/1 to: T34434 from: 1 to: 2232

 439 ACAGACACGAGTACATGGCTAAGAAGTAGTAGTACCACTACAAAGGAAACGA 390
     ||| |||||| |||||||| :::||||| ::
 698 ThrLeuThrSerGluProSerProSerThrGlnSerSerGlyAlaGl 714

 389 AGTACTGTAGTACACAGGTGTCCCAGACCAGGCAGCAAACAGACTATGC 340
       || :|||||||||||||||::::|||||||||||||||||
 714 nSerThrLeuThrThrProSerProAsnProSerGlnSerThrSerSerL 731

 339 TGCCAGCGAGTACAGCCAAGGAAGTCTTCGAGTGCTGC.ACGGTAAACAGC 291
     || :::||||||||::: :::|||:::||| ::::::::::::
 731 euGluSerThrSerGlyAlaThrThrSerGlySerAlaGlyThr 747

 290 ACTCAAATGGCTCCAAGC............GGAGTACAGCAAGGAAG 256
     |||:: :::|||||| ||| :::||||||
 748 ThrMetThrSerProSerGlnSerSerSerValGlySerSerGlnLySe 764

 255 T.........................CATCAAGCCGCTACACAGTA 236
                                ::::||||||||||||||
 764 rThrSerProAlaAlaSerThrThrSerGlyGluMetThrSerGlnGlyS 781

 235 AACAGCACACGAATGGCTCCGATAACATCGACGCCATTGCCGTTGTTAGT 186
     :: :::||| |||||| |||||| :: :::|||||||||||||
 781 erThrGlnThrProGlySerSerValSerThrSerAlaAlaIleLeuThr 797

 185 CTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGACG 136
     ||:::| :::|||||| :::||| ||| :::||||| ||
 798 SerThrGlnGlnSerValSerThrAsnSerProGlySerThrValThrAr 814

 135 ACCCTCAAAAGGGCAGGGA 117
     ||||||||||||||
 814 gProSerThrValSerGly 820

seq_name: pir1:S27981

seq_documentation_block:
homoserine kinase (EC 2.7.1.39) - Pseudomonas aeruginosa
C;Species: Pseudomonas aeruginosa
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 11-Jun-1999
C;Accession: S27981; S19920
R;Clepet, C.; Borne, F.; Krishnapillai, V.; Baird, C.; Patte, J.C.; Caml, B.
Mol. Microbiol. 6, 3109-3119, 1992
A;Title: Isolation, organization and expression of the Pseudomonas aeruginosa threon1
A;Reference number: S27979; MUID:93086420
A;Accession: S27981
A;Molecule type: DNA
A;Residues: 1-316 <CLE>
A;Cross-references: EMBL:X65034; NID:g45424; PIDN:CAA46169.1; PID:g45425
C;Genetics:
A;Gene: thrB
A;Map position: 10 min
C;Superfamily: Pseudomonas homoserine kinase
C;Keywords: phosphotransferase; threonine biosynthesis

alignment_scores:
        Quality: 70.50   Length: 153
        Ratio: 1.306   Gaps: 8
Percent Similarity: 35.294   Percent Identity: 26.797

alignment_block:
US-09-540-234-1 x S27981

Align seg 1/1 to: S27981 from: 1 to: 316

 12 CCCTTATGTATGCAGGAGCGCCTAGTGGGT................. 41
    ||||||||||:::|||||||:::||||
 97 ProAlaLeuLeuGlnProArgLeuAlaGlyArgHisGluArgGlnProAs 113

```
355 AGCAAACAGACTATGCTGCCAGCGAGTACAGCAGCAAGGAAGTCTTCGAGTGC 306
    ||||||               ||||||                  :::||  |||
 51 aAlaAsn......TyrCysGlnSerIleGlyGlyThrGlnValPheIleI 66

305 TGCACGGTAAACAGCACTCA......AATGGCTCCAAGCGGGAGTACA 265
    ||:||                     |||||||||||:::       :::
 66 leSerSerThrGluLeuSerTrpLeuThrAspPheThrSerSerSerLeu 82

264 GCAAGGAAGTCATCAACGGCTACACAGTAAACACACCAACGAATGGCTCCG 215
    ||:|||:      ::|||:||||||| ||:||::::  |||::::  |||
 83 AlaGlnProTrpValAlaThrThrArgAsnThrThrAsn.AsnLysTrpT 99

214 ATAACATCGACGCCATTGGCGTTGTTGTTGTAGTCTTGTACAAACTACGGCTTCA 165
    |||  |||||     ||| :::    |||||  ||||||:   |||||||::::
 99 yrAsnThrAspGlySerSerProTyrSerPhePheTrpThrThrGlyGlu 115

164 CCGACGTCTCGAGGGGCATGTTTGAGACGACGACCCTCAAAAGGGCAGGGACT 115
    |||            :::||||   ||||||          :|||||||  |||
116 ProSerLeuAsnGlyAspCysAlaThrPheLysGlyThrGlyLysAlaAlaGl 132

114 GTTTTCCATTATGCCGTGCCAGCTACACAGTAGCCG......TGTCCGGCTT 71
    ::::::   :::||||||      ||  ||:|||                |||
132 yLeuLysAlaValProCysTyrSerIleGlnProAlaValCysLysGlnM 149

 70 GGAGGTGTCTTTGTCCT 54
    :::||||||||||
149 etProAlaLeuCysPro 154
```

seq_name: pir2:JE0132

seq_documentation_block:
glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) chain wml2 - Watermelon
N;Alternate names: ADPglucose pyrophosphorylase chain wml2; ADPglucose synthase chain
C;Species: Citrullus vulgaris S.
C;Date: 10-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 07-May-1999
C;Accession: JE0132
R;Kim, I.J.; Kahng, H.Y.; Chung, W.I.
Biosci. Biotechnol. Biochem. 62, 550-555, 1998
A;Title: Characterization of cDNAs encoding small and large subunits of ADP-glucose p
A;Reference number: JE0131; MUID:98233279
A;Accession: JE0132
A;Molecule type: mRNA
A;Residues: 1-481 <KIM>
C;Superfamily: glucose-1-phosphate adenylyltransferase
C;Keywords: nucleotidyltransferase

alignment_scores:
         Quality: 69.50   Length: 105
         Ratio: 1.580   Gaps: 4
Percent Similarity: 41.905   Percent Identity: 28.571

alignment_block:
US-09-540-234-1/rev x JE0132

Align seg 1/1   from: 1 to: JE0132 from: 1 to: 481

```
337 CCAGCGAGTACAGCAAGGAAGTCTTCGAGTGCTGCACGGTAAACAGCACT 288
    |||  ||||||||||||||         ::   :::                ::|||
  6 ProArgSerThrAlaArgLysLeuThrProGly........ValAlaTy 19

287 CAAATGGCTCCAAGCGGGAGTACAGCAAGGAAGTCATCAAGCGCTACACAG 238
    :::        |||:|||||||||                .........V 29
 19 rSerValLeuMetSerGluIleSerGlu...... 29

237 TAAACAGCACACAGATGGCTCCGATAACATCGACGCCATTGGCGTTGTTA 188
    |||       :::||||||       |:|||||||||||| 
 29 aIlerSerThrLeuGlnAlaProIlePheGluThrProArgAla...... 43

187 GTCTTGTACAAACTACGGCTTCACCGACGTCTCGAGGGGCATGTTTGAGA 138
    :::||||||
 43 . 43
```

```
 42 ......CATCTGAAGGAA...........AGGACAAAGACACCTC 69
      ||| :::|||      |||  |||||||
113 nAlaHisHisCysGlnGluValAlaIleCysSerAlaThrPheThrProP 130

 70 CAAGCCGGAC.......ACGGCTACTGTAGCTGGCACGGCA 103
    ||                      ||||||||||||||         :::
130 roProAlaGlyAlaSerSerAsnAlaArgAlaThrValAlaCysProGly 146

104 TAATGGAAAACAGTCCCTGCCCTTTGAGGGTCGTCTCAAACATGCCCCT 153
    |||:||::     ||||||::: |||||||  ||| ||||||
147 CysTrpAsnLysValProThrSerProArgAspTyrArgAsnArgProAl 163

154 CGAGACGTCGGTGAGAGCCGTAGTTTGTACAGACTAACACGCCAATGGC 203
    ||              |||  ||:||||                 ||||||
163 aProCysSerProArgPro..........................TrpA 171

204 GTCGGATGGTTATCGGAGGAGCCCATTCGTGTCGTGTTACTGTGTAGCGCTTGAT 253
    ||||||      |||  |||||||::: ||:||||||||||||||         |||
171 rgArgSerProArgTrpMetProSerValGlnArgValGlnArgCysArgAla...... 185

254 GACTTCCTTGCTGTACTCCGCGTTGGAGCCCATTTGAGTGCTGTTTACCG......... 301
    ::::::      |||||:   ||||||::: |||||:||||||:        |||:|||
186 ............ProThrCysMetProTh 191

302 .TGCAGCACTCGAAGACTTCCTTGCTGTACTCGCTGGCAGCATAGTCTGT 350
    ||||||||||:::               |||||||||:::
191 rCysSerAla......ThrThrCysCysSerThrGlyValProHis.......... 203

351 TTGCTGCCTGGTCTGGGGACACCTGTACTACACAGTACTTCGTTTCCCTT 400
    ||                  |||        :::|||:::
204 ........................LeuAlaGlyLeuIleAspPheTyrAsnAla 213

401 TGTAGTGGT 409
    |||||||||
214 CysSerGly 216
```

seq_name: pir2:T20620

seq_documentation_block:
hypothetical protein F08H9.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T20620
A;Experimental source: clone F08H9
R;Wild, A.
submitted to the EMBL Data Library, July 1996
A;Reference number: 219300
A;Accession: T20620
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-346 <WIL>
A;Cross-references: EMBL:Z276557; PIDN:CAB01151.1; GSPDB:GN00023; CESP:F08H9.7
A;Experimental source: clone F08H9
C;Genetics:
A;Gene: CESP:F08H9.7
A;Map position: 5
A;Introns: 17/1; 227/2; 272/2

alignment_scores:
         Quality: 69.50   Length: 123
         Ratio: 1.053   Gaps: 5
Percent Similarity: 53.659   Percent Identity: 26.016

alignment_block:
US-09-540-234-1/rev x T20620

Align seg 1/1   to: T20620 from: 1 to: 346

```
393 ACGAAGTACTGTAGTACACAG.........GTGTCCCAGACCAGGC 356
    |||::|||||||             ||| :::::::||
 35 ThrSerTyrCysTyrrThrLysSerAlaGlnTyrPheSerGluAl 51
```

```
137 CGACCCTCAAAAG.....................................GGCAGGG 118
        ||||||::||                                          |||
44  .AspProLysLysIleAlaSerIleIleLeuGlyGlyGlyAlaGlyThrA 60

117 ACTGTTTCCATTATGCCGTGCCAGCTACAGTAGCCGTCCGGCTTGGA 68
        |||||||||        :::       |||    :::||||
60  rgLeuPheProLeuThrSerGlnArgAlaLysProAlaValProIleGly 76

67  GGTGTCTTTGTCCTT 53
        |||  :::  |||
77  GlyCysTyrArgLeu 81
```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 7, 2001, 16:23:36 ; Search time 1388.89 Seconds
                                                (without alignments)
                                                3021.886 Million cell updates/sec

Title: US-09-540-234-1
Perfect score: 444
Sequence: 1 gggtccacgggcccttatgt.........tgtactcgtgtctgttcgag 444

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 472642650 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*

1: gb_est1:*	44: em_esthum10:*
2: gb_est2:*	45: em_esthum11:*
3: gb_est3:*	46: em_esthum12:*
4: gb_est4:*	47: em_esthum13:*
5: gb_est5:*	48: em_esthum14:*
6: gb_est6:*	49: em_esthum15:*
7: gb_est7:*	50: em_esthum16:*
8: gb_est8:*	51: em_esthum17:*
9: gb_est9:*	52: em_esthum18:*
10: gb_est10:*	53: em_esthum19:*
11: gb_est11:*	54: em_esthum20:*
12: gb_est12:*	55: em_esthum21:*
13: gb_est13:*	56: em_esthum22:*
14: gb_est14:*	57: em_esthum23:*
15: gb_est15:*	58: em_esthum24:*
16: gb_est16:*	59: em_esthum25:*
17: gb_est17:*	60: em_esthum26:*
18: gb_est18:*	61: em_esthum27:*
19: gb_est19:*	62: em_esthum28:*
20: gb_est20:*	63: em_estin1:*
21: gb_est21:*	64: em_estin2:*
22: gb_est22:*	65: em_estin3:*
23: gb_est23:*	66: em_estin4:*
24: gb_est24:*	67: em_estin5:*
25: gb_est33:*	68: em_estom1:*
26: gb_est34:*	69: em_estom2:*
27: gb_est35:*	70: em_estov1:*
28: gb_est36:*	71: em_estov2:*
29: gb_est37:*	72: em_estp11:*
30: gb_est38:*	73: em_estp12:*
31: gb_est39:*	74: em_estp13:*
32: gb_est40:*	75: em_estp14:*
33: em_estba:*	76: em_estp15:*
34: em_estfun:*	77: em_estp16:*
35: em_esthum1:*	78: em_estp17:*
36: em_esthum2:*	79: em_estp18:*
37: em_esthum3:*	80: em_estp19:*
38: em_esthum4:*	81: em_estp110:*
39: em_esthum5:*	82: em_estro1:*
40: em_esthum6:*	83: em_estro2:*
41: em_esthum7:*	84: em_estro3:*
42: em_esthum8:*	85: em_estro4:*
43: em_esthum9:*	86: em_estro5:*
	87: em_estro6:*
	88: em_estro7:*
	89: em_estro8:*
	90: em_estro9:*
	91: em_estro10:*
	92: em_estro11:*
	93: em_estro12:*
	94: em_estro13:*
	95: em_estro14:*
	96: em_estro15:*
	97: em_estro16:*
	98: em_estro17:*
	99: em_estro18:*
	100: em_estro19:*
	101: em_estro20:*
	102: gb_est25:*
	103: gb_est26:*
	104: gb_est27:*
	105: gb_est28:*
	106: gb_est29:*
	107: gb_est30:*
	108: gb_est31:*
	109: gb_est32:*
	110: gb_est41:*
	111: gb_est42:*
	112: gb_est43:*
	113: gb_est44:*
	114: gb_est45:*
	115: gb_est46:*
	116: gb_est47:*

```
117:  gb_est48:.*          190:  gb_est110:.*
118:  gb_est49:.*          191:  gb_est111:.*
119:  gb_est50:.*          192:  gb_htc:.*
120:  gb_est51:.*          193:  em_gss_fun:.*
121:  gb_est52:.*          194:  em_gss_hum1:.*
122:  gb_est53:.*          195:  em_gss_hum2:.*
123:  gb_est54:.*          196:  em_gss_hum3:.*
124:  gb_est55:.*          197:  em_gss_hum4:.*
125:  gb_est56:.*          198:  em_gss_hum5:.*
126:  gb_est57:.*          199:  em_gss_hum6:.*
127:  gb_est58:.*          200:  em_gss_hum7:.*
128:  gb_est59:.*          201:  em_gss_hum8:.*
129:  gb_est60:.*          202:  em_gss_hum9:.*
130:  gb_est61:.*          203:  em_gss_inv1:.*
131:  gb_est62:.*          204:  em_gss_inv2:.*
132:  gb_est63:.*          205:  em_gss_inv3:.*
133:  gb_est64:.*          206:  em_gss_other:.*
134:  gb_est65:.*          207:  em_gss_pln1:.*
135:  gb_est66:.*          208:  em_gss_pln2:.*
136:  gb_est75:.*          209:  em_gss_pro:.*
137:  gb_est76:.*          210:  em_gss_rod1:.*
138:  gb_est77:.*          211:  em_gss_rod2:.*
139:  gb_est78:.*          212:  em_gss_rod3:.*
140:  gb_est79:.*          213:  em_gss_rod4:.*
141:  gb_est80:.*          214:  em_gss_rod5:.*
142:  gb_est81:.*          215:  em_gss_vrt1:.*
143:  gb_est82:.*          216:  em_gss_vrt2:.*
144:  gb_est83:.*          217:  em_gss_vrt3:.*
145:  gb_est84:.*          218:  gb_gss1:.*
146:  gb_est85:.*          219:  gb_gss2:.*
147:  gb_est86:.*          220:  gb_gss3:.*
148:  gb_est87:.*          221:  gb_gss4:.*
149:  gb_est88:.*          222:  gb_gss5:.*
150:  gb_est89:.*          223:  gb_gss6:.*
151:  gb_est90:.*          224:  gb_gss7:.*
152:  gb_est99:.*          225:  gb_gss8:.*
153:  gb_est100:.*         226:  gb_gss9:.*
154:  gb_est101:.*         227:  gb_gss10:.*
155:  gb_est102:.*         228:  gb_gss11:.*
156:  gb_est103:.*         229:  gb_gss12:.*
157:  gb_est104:.*         230:  gb_gss13:.*
158:  gb_est105:.*         231:  gb_gss14:.*
159:  gb_est106:.*         232:  gb_gss15:.*
160:  gb_est67:.*          233:  gb_gss16:.*
161:  gb_est68:.*          234:  gb_gss17:.*
162:  gb_est69:.*          235:  gb_gss18:.*
163:  gb_est70:.*          236:  gb_gss19:.*
164:  gb_est71:.*          237:  gb_gss20:.*
165:  gb_est72:.*          238:  gb_gss21:.*
166:  gb_est73:.*          239:  gb_gss22:.*
167:  gb_est74:.*          240:  gb_gss23:.*
168:  gb_est91:.*          241:  gb_gss24:.*
169:  gb_est92:.*          242:  gb_gss25:.*
170:  gb_est93:.*          243:  gb_gss26:.*
171:  gb_est94:.*          244:  gb_gss27:.*
172:  gb_est95:.*          245:  gb_gss28:.*
173:  gb_est96:.*          246:  gb_gss29:.*
174:  gb_est97:.*          247:  gb_gss30:.*
175:  gb_est98:.*          248:  gb_gss31:.*
176:  em_esthum29:.*       249:  gb_gss32:.*
177:  em_esthum30:.*       250:  gb_gss33:.*
178:  em_esthum31:.*       251:  gb_gss34:.*
179:  em_esthum32:.*       252:  em_gss_inv4:.*
180:  em_esthum33:.*       253:  em_gss_rod6:.*
181:  em_estom3:.*         254:  em_gss_rod7:.*
182:  em_estpln1:.*        255:  em_gss_rod8:.*
183:  em_estro21:.*        256:  gb_gss35:.*
184:  em_estro22:.*        257:  gb_gss36:.*
185:  em_estro23:.*        258:  gb_gss37:.*
186:  em_htc:.*
187:  gb_est107:.*
188:  gb_est108:.*
189:  gb_est109:.*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	289	65.1	572	149	BF482275	BF482275	WHE1798_G
2	237	53.4	880	146	BF264372	BF264372	HV_CEa000
3	102	23.0	271	165	BE230495	BE230495	99AS608 R
4	96.6	21.8	679	165	BE230496	BE230496	99AS609 R
5	88.4	19.9	458	156	C20067	C20067	C20067 Rice
c 6	61.4	13.8	433	108	AU163926	AU163926	AU163926
c 7	61.2	13.8	381	152	BG322351	BG322351	OK-YZ-B15
8	53.2	12.0	482	119	AW677348	AW677348	DG1_5_H04
c 9	43.4	9.8	578	102	AI834587	AI834587	606070D01
c 10	43	9.7	417	154	BG458389	BG458389	947045B04
c 11	43	9.7	750	146	BF267047	BF267047	HV_CEa001
12	38.4	8.6	605	114	AW331079	AW331079	707047A05
c 13	36.8	8.3	581	226	AQ267877	AQ267877	RPCI11-72
14	36.8	8.2	511	137	BE575033	BE575033	946090D12
15	36.6	8.2	578	154	BG458520	BG458520	947045B04
c 16	36.6	8.2	1268	136	BE515313	BE515313	601235904
17	35.6	8.0	613	24	AI770469	AI770469	606053A11
c 18	35	7.9	175	103	AI938961	AI938961	sc63ell.y
19	35	7.9	584	103	AI881530	AI881530	606070D01
20	34.4	7.7	968	219	CNS0062H	CNS0062H	Drosophil
c 21	34.2	7.7	829	221	CNS04AQV	CNS04AQV	Tetraodon
c 22	34	7.7	364	113	AW262166	AW262166	xg62b06.x
23	34	7.7	674	247	AZ632025	AZ632025	1M0486A04
24	33.8	7.6	863	219	CNS013VO	CNS013VO	Drosophil
c 25	33.6	7.6	1101	219	CNS00FEV	CNS00FEV	Drosophil
c 26	33.6	7.5	386	23	AI685592	AI685592	tu20c07.x
27	33.4	7.5	412	140	BE821365	BE821365	GM700024A
c 28	33.4	7.5	523	154	BG508825	BG508825	sac90g12.
c 29	33.4	7.5	535	148	BF424911	BF424911	su52ell.y
c 30	33.4	7.5	544	120	AW733988	AW733988	sk79b06.y
c 31	33.4	7.5	628	242	AZ380524	AZ380524	1M0136D02
c 32	33.4	7.5	933	137	BE563288	BE563288	601335410
c 33	33.4	7.5	945	141	BE871340	BE871340	601449457
c 34	33.2	7.5	287	29	AV327005	AV327005	AV327005
c 35	33.2	7.5	225	23	AI646741	AI646741	ub65c02.x
c 36	33.2	7.5	742	247	AZ640083	AZ640083	1M0502K04
c 37	33	7.4	672	112	AW146599	AW146599	614074C07
c 38	32.8	7.4	416	225	AQ180148	AQ180148	HS_3200_B
c 39	32.8	7.4	715	172	BF974202	BF974202	602243835
c 40	32.8	7.4	887	150	BF575314	BF575314	602134043
c 41	32.8	7.4	1564	153	BG399947	BG399947	602441492
c 42	32.6	7.3	539	231	AQ665985	AQ665985	HS_5351_B
c 43	32.4	7.3	747	240	AZ233398	AZ233398	RPCI-23-8
c 44	32.4	7.3	840	171	BF964711	BF964711	602267814
c 45	32.2	7.3	437	225	AQ175373	AQ175373	HS_3212_B

ALIGNMENTS

RESULT 1

```
BF482275        572 bp   mRNA              EST       06-DEC-2000
LOCUS       WHE1798_G10_N20S Wheat pre-anthesis spike cDNA library Triticum
DEFINITION  aestivum cDNA clone WHE1798_G10_N20, mRNA sequence.
ACCESSION   BF482275
VERSION     BF482275.1  GI:11565576
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 572)
AUTHORS     Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
```
```
TITLE       The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
JOURNAL     Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105599773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1..572
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1798_G10_N20"
                     /clone_lib="Wheat pre-anthesis spike cDNA library"
                     /tissue_type="Spike before anthesis"
                     /dev_stage="Adult plant"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse. Whole spike with awns trimmed, white, green
                     and yellow anther were collected and total RNA, and
                     poly(A) RNA were prepared, a cDNA library was made, and
                     the cDNA  clones were in vivo excised to give pBluescript
                     phagemids in the TJ Close lab (Choi, Close, Fenton) at
                     the University of California, Riverside. Plasmid DNA
                     preparations and DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
BASE COUNT    139 a    134 c     176 g    123 t
ORIGIN
```
```
Query Match             65.1%;   Score 289;  DB 149;   Length 572;
Best Local Similarity   94.4%;   Pred. No. 7e-81;
Matches  322;   Conservative   0;   Mismatches  15;   Indels   4;   Gaps   2;

Qy       6 cacggcgccctatgtatgc---agagcgcctagtggtcatctgaaggaaggacaaag   62
           ||||||| |||||  ||  |    ||| ||||||||||||||||||||| ||||||||
Db     233 CACGGGCCCCTTAGGTTGCCGGCAGCAGCACCTAGTGGTCATCTGAAGGAAGGACAAAG  292

Qy      63 acacctccaagccggacacggctactgtagctggcacggcataatggaaaacagtccctg  122
           ||||||||||||||||||||||  ||||| ||| ||||||||||||| |||||||||||||
Db     293 ACACCTCCAAGCCGGACGCGGCCTAGCTGCTGGCACGGCCATGATGGAAAACAGTCCCTG  352

Qy     123 ccctttgagggtcgtctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtac  182
           ||||||||||||||||||||| ||||| ||||||||||||||||||||||||| ||||||
Db     353 -CCTTTGAGGGTCGTCTCAAATACCGCCCTCGAGACGTCGGTGAAGCCGTAGTTTGTAC  411

Qy     183 aagactaacaacgccaatggcgtcgtgtgtatcggagccattcgtgtctgtgttactgtg  242
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     412 AAGACTAACAACGCCAATGGCGTCGTGTGTATCGGAGCCATTCGTGTCTGTGTTTACTGTG  471

Qy     243 tagcgcttgatgacttcctcgctgctactccgctggagccattgagtgctgttaccgt  302
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     472 TAGCGCTTGATGACTTCCTTGCTGCTACTCCGCTGGAGCCATTGAGTGCTGTTTACCGT  531

Qy     303 gcagcactcgaagacttccttgctgctactgctggcagcat  343
           ||||||||||||||||||||||||||||||||||||||||||
Db     532 GTAGCACTCCGAGTAGTTCCTTGCTGCTACTCCTGCAGCAT  572
```

RESULT 2

```
BF264372        880 bp   mRNA              EST       09-MAR-2001
LOCUS       HV_CEa0009D23f Hordeum vulgare seedling green leaf EST library
DEFINITION  HVcDNA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
            HV_CEa0009D23f, mRNA sequence.
ACCESSION   BF264372
```

```
VERSION      BF264372.2  GI:13261453
KEYWORDS     EST.
SOURCE       barley.
ORGANISM     Hordeum vulgare
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
             ; Triticeae; Hordeum.
REFERENCE    1 (bases 1 to 880)
  AUTHORS    Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
             ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
             ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
             Wood,T.
  TITLE      Development of a genetically and physically anchored EST resource
             for barley genomics
  JOURNAL    Unpublished (2000)
COMMENT      On Nov 17, 2000 this sequence version replaced gi:11195366.
             Contact: Wing RA
             Clemson University Genomics Institute
             Clemson University
             100 Jordan Hall, Clemson, SC 29634, USA
             Tel: 864 656 7288
             Fax: 864 656 4293
             Email: rwing@clemson.edu
             Seq primer: AATTAACCCTCACTAAAGGG
             High quality sequence stop: 819.
FEATURES             Location/Qualifiers
     source          1..880
                     /organism="Hordeum vulgare"
                     /cultivar="CI16155 (M1a13)"
                     /db_xref="taxon:4513"
                     /clone="HV_CE0009D23f"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVcDNA0004 (Erysiphe infected & control)"
                     /tissue_type="seedling green leaf"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      195 a    232 c    272 g    177 t      4 others
ORIGIN

Query Match       53.4%;  Score 237;  DB 146;  Length 880;
Best Local Similarity 82.3%;  Pred. No. 2.6e-64;
Matches 320;  Conservative 0;  Mismatches 59;  Indels 10;  Gaps 4;

Qy    6 cacgggcccttatgtatgcagagagcgcctagtgggtcatctgaaggaaaggacaaagaca 65
          ||| || ||||||| | ||||| || || |||||||||||||||||||||||||| |||||
Db  461 CACGAGCCCTTATGTTTGCAGGAGCACCTAGTGGGTCATCTGAAGGAAAGGACAAAGACA 520

Qy   66 cctccaagccggacacggctactgtagctggcacggcataatggaaaacagtccctgccc 125
          ||| |||||||||||||||| ||||||||||||||||||| | |||||||||| ||||||
Db  521 CCTNCAAGCCGGACACGGCGGCGGCCGACTGTAGCTGGCACGGCATGATGGAAAACAGCCCCTGCCT 580

Qy  126 ttttgagggtcgtctcaaacatgcccctcgagacgtcggtgaag------ccgtagtttgt 180
          ||| ||||||||| | |             |||||||||||||||      |||||||| |||
Db  581 TTTGAGGGTCGTCAGATACGCTCGTTCCTCGAGACGTCGGTGAAGCGTACCGTAGTTTGT 640

Qy  181 acaagactaacaacgccaatggcgtcgatgttatcggagccattcgtgctgtctgttactg 240
          |||||||||||  ||||||||| ||| |||||| |||||||||||| ||| ||| | |||
Db  641 ACAAGACTAACACACCAATGGCCGTC-ATGTTGTCGGAGCCATTCCTGTGCTGTTTGCTG 699

Qy  241 tgtagcgcttgatgacttccttgctactccgcctggagccatttgagtgctgttacc 300
          ||||  || |||||||||  || | |||||| ||||||||||||||||||||||||  ||
Db  700 TGTAATGCTGATGACTTCNTGNGCTGTACTCCTGCTGGGAGCCATTTGAGTGCTGTTA-C 758

Qy  301 gtgcagcactcgaagacttccttgctgctactccgctgg---cagcatagtctgttgctgc 357
          |||  | || |||||||||  ||  ||| |||||||||||   ||| |||| |||| |||||
Db  759 GTGGTACACTGGAAGACTTNCTTACTGTACTCGCTGGTACTCGCTGGGCAAGCTATCGTCTGGTTGCTGG 818

Qy  358 ctgtctggggacacctgtgtactacagt 386
          |||||| ||||| | ||||| || ||| |
Db  819 GTGGGCTGGGAACACTGTGTCCAAACGT 847
```

```
RESULT 3
BE230495
LOCUS        BE230495          271 bp    mRNA                    EST    07-JUL-2000
DEFINITION   99AS608 Rice Seedling Lambda ZAPII cDNA Library Oryza sativa cDNA
             clone 99AS608, mRNA sequence.
ACCESSION    BE230495
VERSION      BE230495.1  GI:8956580
KEYWORDS     EST.
SOURCE       Oryza sativa.
ORGANISM     Oryza sativa
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1 (bases 1 to 271)
  AUTHORS    Lee,M.C., Shin,Y.C., Lee,T.H., Jeong,S.H., Kim,J.K., Eun,M.Y. and
             Nahm,B.H.
  TITLE      Large-scale Sequencing Analysis of ESTs from Rice Seedling
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Eun M.Y.
             Department of Cytogenetics
             National Inst. of Agri. Sci. and Tech, RDA
             Suwon, Kyunggido, Korea
             Tel: 82 331 290 0301
             Fax: 82 331 290 0307
             Email: myeun@sun20.asti.re.kr.
FEATURES             Location/Qualifiers
     source          1..271
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /db_xref="taxon:4530"
                     /clone="99AS608"
                     /clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector at 5'end with EcoRI and 3' end with Xho I site"
BASE COUNT       66 a     66 c     72 g     66 t      1 others
ORIGIN

Query Match       23.0%;  Score 102;  DB 165;  Length 271;
Best Local Similarity 72.6%;  Pred. No. 1.4e-21;
Matches 146;  Conservative 0;  Mismatches 51;  Indels 4;  Gaps 1;

Qy    6 cacgggcccttatgtatgcagagagcgcctagtgggtcatctgaaggaaaggacaaagaca 65
          ||| || | ||| | ||||| |  ||| || || ||| |||| | || |||  |||||||||
Db   12 CACGAGCTCTTAGGTTTGCAGAAACATCTAGTGGACCATCTCAAGAAAAATGGCAAAGACA 71

Qy   66 cctccaagccggacacggctactgtagctggcacggcataatggaaaacagtccctgccc 125
          ||| ||||||||||| |||  || || | || ||||||||  | ||    ||| |||||||
Db   72 GCTCCAAGCCGGACCGGCCGCCGCCACTGTGCGAGGCACGGCGACGGCGACCACTCCCTGCCTT 131

Qy  126 tttgagggtcgtctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaag 185
          || |||   |||||  |||| ||| || | | ||| ||||||||||||||||||||||| |
Db  132 TTGTGAG---ATCTCGGACACATGGCCCCTTACTTTCGGGGTGAAGCCGTAGTTTGTACAAC 187

Qy  186 actaacaacgccaatggcgtc 206
          ||||||  | ||| ||  | ||
Db  188 ACTAACGACAATCAGNGGCATC 208
```

```
RESULT 4
BE230496
LOCUS        BE230496         .679 bp    mRNA                    EST    07-JUL-2000
DEFINITION   99AS609 Rice Seedling Lambda ZAPII cDNA Library Oryza sativa cDNA
             clone 99AS609, mRNA sequence.
ACCESSION    BE230496
VERSION      BE230496.1  GI:8956581
KEYWORDS     EST.
SOURCE       Oryza sativa.
ORGANISM     Oryza sativa
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
REFERENCE    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Ehrhartoideae; Oryzeae; Oryza.
AUTHORS      Lee,M.C., Shin,Y.C., Lee,T.H., Jeong,S.H., Kim,J.K., Eun,M.Y. and
             Nahm,B.H.
TITLE        Large-scale Sequencing Analysis of ESTs from Rice Seedling
JOURNAL      Unpublished (1999)
COMMENT      Contact: Eun,M.Y.
             Department of Cytogenetics
             National Inst. of Agri. Sci. and Tech, RDA
             Suwon, Kyunggido, Korea
             Tel: 82 331 290 0301
             Fax: 82 331 290 0307
             Email: myeun@sun20.asti.re.kr.
FEATURES           Location/Qualifiers
     source        1..679
                   /organism="Oryza sativa"
                   /cultivar="Milyang23"
                   /db_xref="taxon:4530"
                   /clone="99AS609"
                   /clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
                   /dev_stage="5 days after pollination"
                   /lab_host="E. coli SOLR"
                   /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                   XhoI; Directional cDNA library inserted into lambda ZAPII
                   vector at 5' end with EcoRI and 3' end with Xho I site"
BASE COUNT   179 a      171 c      176 g      150 t       3 others
ORIGIN

Query Match         21.8%;     Score 96.6;  DB 165;  Length 679;
Best Local Similarity   71.1%;  Pred. No. 9.4e-20;
Matches 143;   Conservative    0;  Mismatches   54;  Indels    4;  Gaps     1;

Qy      6 cacgggcccttatgtatgcaggagcgcctagtgggtcatctgaaggaaaggacaaagaca 65
          |||||| |||||| || |||||| || ||||||| || ||| || |||| |||||||||
Db     12 CACGAGCTCTTAGGTTTGCAGAAACATCTAGTGGACCATCTCAAGAAAATGGCAAAGACA 71

Qy     66 cctccaagccggacacggctactgtagctggcacggcataatgaaaacagtccctgccc 125
          ||||||||||| || |||| |||||| |||||||| || ||||| || |
Db     72 GCTCCAAGCCGGACGCCGGCCACTGTGGCGAGGCACGGCTTATTGGAAGCACTCCCTGCCTT 131

Qy    126 tttgagggtcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaag 185
          || |||     |||| ||||  |||  ||||||||||||||||||||||||||||
Db    132 TTGTGAG---ATCTCGGACATGGCCCTTACTTTTGGGGTGAAGCCGTAGTTTGTACAAC 187

Qy    186 actaacaacgccaatggcgtc 206
          |||||| |||| |||| |||
Db    188 ACTAACGACATCAGTGGCATC 208

RESULT 5
C20067/c
LOCUS       C20067          458 bp   mRNA              EST        24-OCT-1996
DEFINITION  C20067 Rice panicle at ripening stage Oryza sativa cDNA clone
            E11462_1A, mRNA sequence.
ACCESSION   C20067
VERSION     C20067.1    GI:1632338
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 458)
AUTHORS     Sasaki,T.
TITLE       Rice cDNA from panicle at ripening stage
JOURNAL     Unpublished (1996)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
```

```
                                                          Japan 305
                                                          Tel: 0298-38-7441
                                                          Fax: 0298-38-7468
                                                          Email: tsasaki@abr.affrc.go.jp.
FEATURES           Location/Qualifiers
     source        1..458
                   /organism="Oryza sativa"
                   /strain="Nipponbare"
                   /db_xref="taxon:4530"
                   /clone="E11462_1A"
                   /clone_lib="Rice panicle at ripening stage"
                   /dev_stage="ripening stage"
                   /note="Organ: panicle; Rice cDNA from panicle at ripening
                   stage"
BASE COUNT   124 a      113 c      103 g      117 t       1 others
ORIGIN

Query Match         19.9%;     Score 88.4;  DB 156;  Length 458:
Best Local Similarity   68.5%;  Pred. No. 3.4e-17;
Matches 170;   Conservative    0;  Mismatches   66;  Indels   12;  Gaps     3;

Qy      6 cacgggcccttatgtatgcaggagcgcctagtgggtcatctgaaggaaaggacaaagaca 65
          |||||| |||||| || |||||| || ||||||| || ||| || |||| |||||||||
Db    345 CACGAGCTCTTAGGTTTGCAGAAACATCTAGTGGACCATCTCAAGAAAATGGCAAAGACA 286

Qy     66 cctccaagcc-----ggacacggctactgtagctgcacggcataatggaaaacagtcc 119
          |||||||||      || |||| |||||| ||||||| || ||||| ||||||
Db    285 GCTCCAAGCCGGACGCCGGCCACTGTGGCGAGGCACGGCATAATGGAAGCACTCCC 226

Qy    120 ctgccctttgagggtcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttg 179
          || |||  |||||  ||||  |||  |||  ||||||||||||||||||||||
Db    225 TGCCTTTGTGAG---ATCTCGGACATGGCCCTTACTTTTGGGGTGAAGCCGTAGTTTG 170

Qy    180 tacaagactaacaacgccaatggcgtc--gatgttatcggagccattcgtgtgctgttta 237
          |||||| |||| |||| |||| |||    |||| ||| || ||||| || ||||| |||||
Db    169 TACAACACTAACGACATCAGTGGCATCATGTTTCTCCTAGCCATGTTGAGTGCTATTTA 110

Qy    238 ctgtgtag 245
          |||||||
Db    109 CTGTGTTG 102

RESULT 6
AU163926/c
LOCUS       AU163926           433 bp   mRNA              EST        14-NOV-2000
DEFINITION  AU163926 Rice panicle at ripening stage Oryza sativa cDNA clone
            E11462, mRNA sequence.
ACCESSION   AU163926
VERSION     AU163926.1    GI:11171718
KEYWORDS    EST.
SOURCE      Oryza sativa.
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 433)
AUTHORS     Sasaki,T. and Yamamoto,K.
TITLE       Rice cDNA from panicle at ripening stage (2000)
JOURNAL     Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
            E11462_42.
FEATURES           Location/Qualifiers
     source        1..433
```

```
                                          /organism="Oryza sativa"
                                          /strain="Nipponbare"
                                          /db_xref="taxon:4530"
                                          /clone="E11462"
                                          /clone_lib="Rice panicle at ripening stage"
                                          /dev_stage="ripening stage"
                                          /note="Organ: panicle; Rice cDNA from panicle at ripening
                                          stage"
BASE COUNT       77 a      138 c      93 g     124 t
ORIGIN

Query Match          13.8%;     Score 61.4;    DB 108;   Length 433;
Best Local Similarity     72.2%;     Pred. No. 1.3e-08;
Matches  96;    Conservative   0;    Mismatches  31;    Indels  6;   Gaps  1;

Qy       6 cacgggccctatgtatgcaggagcgcctagtgggtcatctgaaggaaggacaaagaca 65
           ||||| |||| |||||||||||| | |||   ||||||||||||||||||||||||||
Db     160 CACGAGCTCTTAGGTTTGCAGAAACATCTAGTGGCACCATCTCAAGAAAATGGCAAAGACA 101

Qy      66 cctccaagcc------ggacacggctactgtagtcgcacggcataatggaaaacagtcc 119
           |||||||||       || ||||||||||| |  |||||||||||||||| |||||||||
Db     100 GCTCCAAGCCGACCGAGCGCGGACGCGGCCACTGTGCAGGCACGGCATAATGGAAGCACTCCC 41

Qy     120 ctgcccttgag 132
           | || |||||
Db      40 TGCCTTTGTGAG 28

RESULT  7
BG322351
LOCUS        BG322351      381 bp     mRNA            EST       27-FEB-2001
DEFINITION   OK-YZ-B156 Bermudagrass/SDS SSH Library Cynodon dactylon cDNA clone
             91-86.M13 similar to hypothetical protein F9G14.80 - Arabidopsis
             thaliana gi|7413552|emb|CAB86031.1| (AL162973) putative protein
             [Arabidopsis thaliana], mRNA sequence.
ACCESSION    BG322351
VERSION      BG322351.1  GI:13152029
KEYWORDS     EST.
SOURCE       Bermuda grass.
ORGANISM     Cynodon dactylon
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
             clade; Chloridoideae; Cynodonteae; Cynodon.
REFERENCE    1  (bases 1 to 381)
AUTHORS      Zhang,Y. and Guenzi,A.C.
TITLE        Detection of Fungal-induced cDNAs from Bermudagrass using
             Subtractive Hybridization Coupled to cDNA Microarray
JOURNAL      Unpublished (2001)
COMMENT      Contact: Guenzi AC
             Dep. of Plant and Soil Sciences
             Oklahoma State University
             368 Agriculture Hall, Stillwater, OK 74078-6028, USA
             Tel: 405 744 5532
             Fax: 405 744 6039
             Email: acg@okstate.edu
             Research supported by USGA grant 1998-11-119.
             PCR Primers
             FORWARD: M13 Forward
             BACKWARD: M13 Reverse
             Seq primer: M13 Reverse
             High quality sequence stop: 381.
FEATURES             Location/Qualifiers
     source          1..381
                     /organism="Cynodon dactylon"
                     /cultivar="Jackpot"
                     /db_xref="taxon:28909"
                     /clone="91-86.M13"
                     /clone_lib="Bermudagrass/SDS SSH Library"
                     /tissue_type="Crown"
                     /lab_host="E.coli"
                     /note="Vector: pCR 2.1-TOPO; Site_1: EcoR I; Site_2: EcoR
                     I; Clontech PCR-select cDNA subtraction library"
```

```
BASE COUNT       99 a       86 c      106 g      90 t
ORIGIN

Query Match          13.8%;     Score 61.2;    DB 152;   Length 381;
Best Local Similarity     73.6%;     Pred. No. 1.4e-08;
Matches  78;    Conservative   0;    Mismatches  28;    Indels  0;   Gaps  0;

Qy       6 cacgggccctatgtatgcaggagcgcctagtgggtcatctgaaggaaggacaaagaca 65
           ||||||||  |||||||||| |||||||||||||||||||||||||||||||||||||
Db      58 CACGGGCCCAGAGGTTTGCTGAAACACCTAGTGGATCACCCCAGGAAATGGCAAAGACA 117

Qy      66 cctccaagccggacacggctactgtagtcgcacggcataatggaa 111
           ||| |||||||| ||||||| |||| ||||| |||||||| ||| |||
Db     118 ACTCGAAGCCGGATGACGCCGCTGTGACAGGCACAGCATGATCGAA 163

RESULT  8
AW677348
LOCUS        AW677348      482 bp     mRNA            EST       19-JUL-2000
DEFINITION   DG1_5_H04.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA
             sequence.
ACCESSION    AW677348
VERSION      AW677348.1  GI:7550928
KEYWORDS     EST.
SOURCE       sorghum.
ORGANISM     Sorghum bicolor
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
             clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE    1  (bases 1 to 482)
AUTHORS      Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
             ,L.H.
TITLE        An EST database from Sorghum: dark-grown seedlings
JOURNAL      Unpublished (2000)
COMMENT      Contact: Cordonnier-Pratt MM
             Department of Botany
             The University of Georgia
             Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
             Tel: 706 542 1860
             Fax: 706 542 1805
             Email: mmpratt@uga.edu
             Sequences have been trimmed to exclude PolyA, vector and regions
             below Phred quality 16. The threshold for highest quality sequence
             is 20.
             Seq primer: T7
             High quality sequence start: 2
             High quality sequence stop: 482
             POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..482
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Dark Grown 1 (DG1)"
                     /note="Organ: 5-day-old dark-grown seedlings; Vector:
                     Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was
                     made from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      114 a      100 c      130 g      138 t
ORIGIN

Query Match          12.0%;     Score 53.2;    DB 119;   Length 482;
Best Local Similarity     56.1%;     Pred. No. 5.2e-06;
Matches 165;    Conservative   0;    Mismatches 118;    Indels  11;   Gaps  3;

Qy       6 cacgggccctatgtatgcaggagcgcctagtgggtcatctgaaggaaggacaaagaca 65
           ||||||| | ||||||||| ||| |||||||||||||||||||||||||||||||||||
Db      53 CCCGAGCAGCAAGATTTGGAGATACATCTAGTGGATCAGCTCAGGAAAATGGCAAAGACA 112

Qy      66 cctccaagccggacacggctactgtagtcgcacggcataatggaaaaacagtccctgcc 125
           ||| ||||||| ||||||| |||| ||||| |||||||| ||||||||||||  |||||
Db     113 GCTCGAAGCTGGACGCAGCTGCTGTTACAGGCACGCAGCTTGATTCAG------CAGTCCG 166
```

Qy   126 tttgagggtcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaag 185
                ||| | ||||||| | | || | | | ||||||||| | ||||||||
Db   167 GCTTTGATGAGGTCCAGATATATCCGACTT---TTGGTGAAGCCATAGCTTGTACACC 223

Qy   186 actaacaacgccaatggcgtcga--tgttatcggagccattcgtgctgtttactgtgt 243
                ||||||| || | | -| |||| | ||||||||| ||||||
Db   224 ACTAACTGCATGAACGGCCATCACGTGTTTCCCAGTCATCGCAGTGTGATGTTTACCGTGT 283

Qy   244 agcgcttgatgacttcctgctgctactccgcttggagccattggagtgctgttt 297
                | |||||||| | | | ||| ||||||| |||| || ||
Db   284 ATGACTTCTTAGTTGCTGGCACTGGCACTGGCCTAGATTTGCTGCCTGGTT 337

RESULT 9
AI834587/c
LOCUS   AI834587   578 bp   mRNA   EST   02-FEB-2000
DEFINITION 606070D01.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI834587
VERSION   AI834587.1 GI:568796
KEYWORDS   EST.
SOURCE   Zea mays.
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 578)
AUTHORS   Walbot,V.
TITLE   Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606070 row: D column: 01.
FEATURES   Location/Qualifiers
     source   1..578
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XIOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT   149 a   152 c   124 g   151 t   2 others
ORIGIN

Query Match   9.8%;   Score 43.4;   DB 102;   Length 578;
Best Local Similarity   57.5%;   Pred. No. 0.0071;
Matches 157;   Conservative   0;   Mismatches 106;   Indels   10;   Gaps   4;

Qy   20   tatgcaggagcgcctagtgggt---catctgaaggaaggacaaagacacctccaagccg 76
                ||||||||| |||||| | | ||||||||||||| |||| ||||||||||
Db   480 TTTGCAGAAGCGTCTGGTGGATCAGCAGCGTCGGGAAAACGGCAAAGACAGCTCGAAGCCG 421

Qy   77   gacacggctactgtagctgcacggcataatggaaaacagtccctgcccttttgagggtc 136
                |||||||| | |||||||| || |||| | | |||||| || |||||||||| |
Db   420 GACGCAGCCGCCGTGTTACCGGCACCAGCTTGATTGAG-CACTATGCCTTGACGAGGTCTC 362

Qy   137 gtctcaaacatgccctcgagacgtcggtgaagccgtag--tttgtacaagactaacaac 194
                |||||||| |||||| |||||||||| |||||||| | |||||||||||| ||
Db   361 AGAAATATTATCTCCTAGGAGACTTTGGTGAAGCCATAGCTTTGTACACCACTAACCGC 302

Qy   195 gccaatggctcga---tgttatcggagccattcgtgctgttcactgtgtagcgctt 250
                | |||| | | ||||| | || |||||| |||| | | ||||| |||||
Db   301 ATGGACGGGGGGCCTCGTGTTTCCCAGTCATCGTTCAGTCATGTTCAGTGATGTGTTACCGGCTATGACTT 242

Qy   251 gatgacttccttgctgtactccgcttggagcca 283
                || | ||||||| || | |||||||| ||
Db   241 TATTAGTTGCTGGGCACACTGGCAGTGTAGCTA 209

RESULT 10
BG458389/c
LOCUS   BG458389   417 bp   mRNA   EST   19-MAR-2001
DEFINITION 947045B04.x2 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
            sequence.
ACCESSION   BG458389
VERSION   BG458389.1 GI:13381609
KEYWORDS   EST.
SOURCE   Zea mays.
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 417)
AUTHORS   Walbot,V.
TITLE   Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 947045 row: B column: 04.
FEATURES   Location/Qualifiers
     source   1..417
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="947 - 2 week shoot from Barkan lab"
                     /tissue_type="leaf and stem, including leaf base"
                     /dev_stage="2 week old seedling (3 leaves)"
                     /lab_host="XL1-Blue"
                     /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);
                     Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
                     Stratagene's UniZap XR cDNA cloning kit with the 5' end
                     at the EcoRI site. The library represents 8 x 10e5
                     independent recombinant phage. The plants were greenhouse
                     grown."
BASE COUNT   100 a   118 c   85 g   113 t   1 others
ORIGIN

Query Match   9.7%;   Score 43;   DB 154;   Length 417;
Best Local Similarity   53.7%;   Pred. No. 0.0087;
Matches 159;   Conservative   0;   Mismatches 126;   Indels   11;   Gaps   3;

Qy   20   tatgcaggagcgcctagtgggt---catctgaaggaaggacaaagacacctccaagccg 76
                ||||||||| |||||| | | ||||||||||||| |||| ||||||||||
Db   310 TTTGCAGAAGCGTCTGGTGGATCAGCAGCGTCGGGAAAACGGCAAAGACAGCTCGAAGCCG 251

Qy   77   gacacggctactgtagctgcacggcataatggaaaacagtccctgcccttttgagggtc 136
                |||||||| | |||||||| || |||| | | |||||| || |||||||||| |
Db   250 GACGCAGCCGCCGTGTTACCGGCACCAGCTTGATTGAG-CACTATGCCTTGACGAGGTCTC 192

Qy   137 gtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaagactaacaacgc 196
                |||||||| |||||| |||||||||| |||||||| |||||||||| || |
Db   191 AGAAATATTATCTCCTAGGAGACTTTGGTGAAGCCATAGCTTTGTGAAGAAAAAAAAAAA 132

Qy   197 caatggcgtcga------tgttatcggagccattcgtgctgttactgtgtagcgct 249
                | |||| | | ||||| | || |||||| |||| |||| ||||| |||||
Db   131 CATGGACGGGGGGCGTCCGTGTTTCCCAGTCATCGTTCAGTGATGTTTACCGGCTATGACT 72

```
Qy    250  tgatgacttcctgctgtactccgcttgagccatttgagtgctgtttaccgtgca  305
                ||| | || ||||| |  ||||||   |  ||||  |||||   ||| |||||
Db     71  TTATTAGTTGCTGGGCACACTGGCAGTGTAGCTAGATTTGCTGTGTGGTTTGTGGA   16
```

RESULT 11
BF267047
LOCUS   BF267047   750 bp   mRNA   EST   09-MAR-2001
DEFINITION HV_CEa0016M17f Hordeum vulgare seedling green leaf EST library
            HVCDNA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
            HV_CEa0016M17f, mRNA sequence.
ACCESSION   BF267047
VERSION   BF267047.2 GI:13263140
KEYWORDS   EST.
SOURCE   barley.
ORGANISM   Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 750)
AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
TITLE   Development of a genetically and physically anchored EST resource
            for barley genomics
JOURNAL   Unpublished (2000)
COMMENT   On Nov 17, 2000 this sequence version replaced gi:11198042.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 745.
FEATURES   Location/Qualifiers
     source   1..750
                     /organism="Hordeum vulgare"
                     /cultivar="CI16155 (Mla13)"
                     /db_xref="taxon:4513"
                     /clone_lib="HV_CEa0016M17f"
                     /clone="Hordeum vulgare seedling green leaf EST
                     library HVCDNA0004 (Erysiphe infected & control)"
                     /tissue_type="seedling green leaf"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT   177 a   218 c   227 g   128 t
ORIGIN

Query Match   9.7%; Score 43; DB 146; Length 750;
Best Local Similarity 65.5%; Pred. No. 0.01;
Matches   78; Conservative   0; Mismatches 40; Indels 1; Gaps 1;

Qy   8 cgggccctatgtatgcaggaggcgcctagtggtcatctgaaggaaaggacaaagacacc   67
              || || | |||||||| ||||||| ||| || |||| |||||||||||||| ||||
Db   485 CCGAGCCCTTAGTTGTCAGGAGGACCCCCTA-TGGGGCCTTCTGTAGGAAAGGACAAAAACACC 543

Qy   68 tccaagccggacacggctactgtagctggcacggcataatggaaaacagtccctgccct 126
              | ||||||| ||||| ||||||||| ||||||||||||| |||||| ||||||||
Db   544 TGCAACCCGGACCCCGTAACTGATCTCGGTCCCCGTATTGAAGAAAAACCACCCCGT 602

RESULT 12
AW331079
LOCUS   AW331079   605 bp   mRNA   EST   31-JAN-2000
DEFINITION 707047A05.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea
            mays cDNA, mRNA sequence.
ACCESSION   AW331079
VERSION   AW331079.1 GI:6827436
```

---

KEYWORDS    EST.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 605)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 707047 row: A column: 05.
FEATURES             Location/Qualifiers
     source          1..605
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="707 - Mixed adult tissues from Walbot lab (SK
                     )"
                     /tissue_type="tassel, kernel, silk, husk, root, leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: tassel, kernel, silk, husk, root, leaf;
                     Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
                     differentiated maize tissues from an active Mutator
                     plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
                     husk, root, leaf).   Unidirectionally cloned."
BASE COUNT   154 a    155 c    191 g    105 t
ORIGIN

Query Match         8.6%;  Score 38.4;  DB 114;  Length 605;
Best Local Similarity  58.3%;  Pred. No. 0.28;
Matches  105;  Conservative   0;  Mismatches  71;  Indels  4;  Gaps  2;

Qy     20  tatgcaggaggcgcctagtggt---catctgaaggaaaggacaaagacacctccaagccg   76
              |||||||||| |||| ||||||   ||| ||||||||||||||||||||||||||||||
Db    422  TTTGCAGAAGCGTCTGGTGGATCAGCAGCTCGGAAAACGGCAAAGACAGCTCCAAGCCG  481

Qy     77  gacacggctactgtagctggcacggcataatggaaaacagtccctgcccctttgagggtc  136
              ||||| ||||||||||||||||| |||||||||||||||||||| ||||||||||||||
Db    482  GACGCAGCCGCCGCTTACCGGCACAGCTTGATTTGAG-CACTCTGCCTGCGGCGAGGTCTC  540

Qy    137  gtctcaaacatgccccctgagacgtcggtgaagccgtagtttgtacaagactaacaacgc  196
              |  |||||||||||  ||||||||| |||| | || || ||| |||||||||| |  ||
Db    541  AGATATATTGTCTCCTAGGAGACTTTGTGAAGCCATAGCTTTTGTACACCTAACCGC  600

RESULT 13
AQ267877/c
LOCUS       AQ267877    581 bp    DNA          GSS        27-APR-1999
DEFINITION  RPCI11-72E14.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-72E14,
            DNA sequence.
ACCESSION   AQ267877
VERSION     AQ267877.1  GI:3795481
KEYWORDS    GSS.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 581)
AUTHORS     Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE       Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL     Unpublished (1998)
COMMENT     Other_GSSs: RPCI11-72E14.TK

```
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@djong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES         Location/Qualifiers
   source        1..581
                 /organism="Homo sapiens"
                 /db_xref="GDB:7527373"
                 /db_xref="taxon:9606"
                 /clone="RPCI-11-72E14"
                 /clone_lib="RPCI-11"
                 /sex="Male"
                 /cell_type="lymphocytes"
                 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                 RPCI11 Human Male BAC Library"
BASE COUNT     260 a    129 c    81 g    111 t
ORIGIN

Query Match         8.3%;     Score 36.8; DB 226;  Length 581;
Best Local Similarity  48.1%;  Pred. No. 0.88;
Matches 104;  Conservative  0;  Mismatches 112;  Indels 0;  Gaps 0;

Qy    200 tggcgtcgatgtatcggagccattcgtgtcgtttactgtgtagccgcttgatgacttc 259
              || ||| |||| | ||||| || | || || | ||||||| | |||||
Db    217 TGTTGTCTATGATTTTTCAGCAGTGTTTTCGTTTTTCTTGTGAGGTATTTCACTTC 158

Qy    260 cttgctgactccgcttggagccattgagtgctgttaccgtgcagcactcgaagagactt 319
              || | | | ||| ||| | | |||| || | | ||||| | |
Db    157 CTTGGTTAGGTATATTCCTAAGCATTTTATTTTATTTTATTTTTGCAGCTGTTGTAAAAGG 98

Qy    320 ccttgctgactccgctgcagcatagtctgtttgctgcctggtctgggacacctgtgta 379
              ||| | || | | |||| | | |||||| || | ||||
Db     97 GGTTGAGTTCTTGACTTGCATTCTCAGCTTGGTTGCTGGTGGTGTATAGGAGAGCTACTGA 38

Qy    380 ctacagtactcgtttcccttgtagtggtactact 415
              | |||| | |||||||| |||
Db     37 TTTGCGTACATTAATTTTGTGTCTGGAAACGTTACT 2

RESULT 14
BE575033
LOCUS       BE575033       511 bp      mRNA        EST       15-AUG-2000
DEFINITION  946090D12.y1 946 - tassel primordium prepared by Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BE575033
VERSION     BE575033.1  GI:9824936
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 511)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
```

```
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 946090 row: D column: 12.
FEATURES         Location/Qualifiers
   source        1..511
                 /organism="Zea mays"
                 /cultivar="OH43"
                 /db_xref="taxon:4577"
                 /clone_lib="946 - tassel primordium prepared by Schmidt
                 lab"
                 /tissue_type="tassels"
                 /dev_stage="just after the transition from vegetative to
                 inflorescence development"
                 /lab_host="XLOLR"
                 /note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
                 Site_2: XhoI; George Chuck dissected immature tassels
                 between 1mm and 3mm.  Sharon Stanfield prepared the cDNA
                 library in HybriZAP.  Sample insert size range was 350 bp
                 to 3 Kb with a 1 Kb average."
BASE COUNT     130 a    132 c    169 g    80 t
ORIGIN

Query Match         8.2%;     Score 36.6; DB 137;  Length 511;
Best Local Similarity  70.3%;  Pred. No. 0.99;
Matches 64;  Conservative  0;  Mismatches 24;  Indels 3;  Gaps 1;

Qy     20 tatgcaggagcgcctagtgggt---catctgaaggaaggacaaagacacctccaagccg 76
              |||||| |||||||||| ||| |||||||| | |||| | |||||
Db    398 TTTGCAGAAGCGCTCTCGGTGGATCAGCAGCTCGGGAAAACGGCAAAGACAGCTCGAAGCCG 457

Qy     77 gacacggctactgtagctgcacgcgcatat 107
              || ||| ||||| | || ||| | ||||
Db    458 GACCGCAGCCGCCGTGTTACCGGCACACGTTGAT 488

RESULT 15
BG458520
LOCUS       BG458520       578 bp      mRNA        EST       19-MAR-2001
DEFINITION  947045B04.y1 947 - 2 week shoot from Barkan lab zea mays cDNA, mRNA
            sequence.
ACCESSION   BG458520
VERSION     BG458520.1  GI:13381845
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 578)
AUTHORS     Walbot,V.
TITLE       Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL     Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 947045 row: B column: 04.
FEATURES         Location/Qualifiers
   source        1..578
                 /organism="Zea mays"
                 /cultivar="B73"
                 /db_xref="taxon:4577"
                 /clone_lib="947 - 2 week shoot from Barkan lab"
                 /tissue_type="leaf and stem, including leaf base"
                 /dev_stage="2 week old seedling (3 leaves)"
                 /lab_host="XL1-Blue"
                 /note="Organ: shoot; Vector: Lambda ZAP (pBlueScript SK-);
```

```
Qy    209 tgttatcggaggcattcgtgtgctgttacttgtgtagcgcttgatgacttccttgcgta 268
          |||| || ||||| |||| ||||||| || || || ||| |||||| ||   ||||
Db    712 TGCTCTGTGTGCCCCGGTGTGTGCGTGTTGTTGTTGCGTGTGGCCTTGTTGTTTCCTTT 653

Qy    269 ctccgctggagccattgagtgctgtttaccgtgcagcactcgaagacttccttgctgt 328
          ||  ||||| |||||||  || ||| || ||||||||||| |||||||| ||||||
Db    652 GTCCTGTCTGTGTGTGCGCTCGCGCCGTGTGCGCGCCGCGTTGCTTGTGTGTCTTGCTCT 593

Qy    329 actcgtggcagcatagtctgtttgtgctgtctcgggacaacctgtgtactacagtac 388
          ||||| ||| || |||| ||||| ||| || ||||| |||| ||||| || ||
Db    592 TCTGTCTCTCGGCTTGCTTCCGGCTGCTTGTGTGGTGGTGGCTGCCCTGCCTGCTGTTT 533

Qy    389 ttcgttcccttgt 403
          || || ||||||
Db    532 GTTGTCTCCCTTGT 518
```

RESULT 17
AI770469
LOCUS   AI770469   613 bp   mRNA   EST   02-FEB-2000
DEFINITION 60605A11.x2 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI770469
VERSION   AI770469.1   GI:5268505
KEYWORDS   EST.
SOURCE   Zea mays.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 613)
AUTHORS   Walbot,V.
TITLE   Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
            Department of Biological sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606053 row: A column: 11.
FEATURES   Location/Qualifiers
     source   1..613
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT   156 a   169 c   187 g   101 t
ORIGIN

Query Match   8.0%; Score 35.6; DB 24; Length 613;
Best Local Similarity   71.2%; Pred. No. 2.2;
Matches   47; Conservative   0; Mismatches   19; Indels   0; Gaps   0;

Qy   42 catctgaaggaaggacaaagacacctccaagccggacacggctactgagctggcacgg 101
          |||||||| || || ||| || || |||| ||| ||||| || ||| ||||| |||
Db   508 CAGCTCGGAAAAACGGCAAAGACAGCTCGAAGCACCGGACGCAGCCGCCGTCGTTACCGGCACAG 567

Qy   102 cataat 107
          || |||
Db   568 CTTGAT 573

```
            Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
            Stratagene's UniZap XR cDNA cloning kit with the 5' end
            at the EcoRI site. The library represents 8 x 10e5
            independent recombinant phage. The plants were greenhouse
            grown."
BASE COUNT     139 a    156 c    182 g    101 t
ORIGIN

Query Match              8.2%;  Score 36.6;  DB 154;  Length 578;
Best Local Similarity   70.3%;  Pred. No. 1;
Matches   64;  Conservative    0;  Mismatches   24;  Indels    3;  Gaps    1;

Qy     20 tatgcaggagcgcctagtgggt---catctgaaggaaggacaaagacacctccaagccg 76
          |||||||| ||||||| ||||||     |||||||| || ||  ||| || || ||||
Db    465 TTTGCAGAAGCGGTCTCGGTGGGATCAGCAGCTCGGGATCAGCAGCTCGGGATCAGCAGCTCGAAGCCG 524

Qy     77 gacacggctactgtgtagctgcacggcataat 107
          ||  |||| |||| |||| ||||| |||| ||
Db    525 GACGCAGCCGCCGTCGTTACCGGCACAGCTTGAT 555
```

RESULT 16
BE515313/c
LOCUS   BE515313   1268 bp   mRNA   EST   07-AUG-2000
DEFINITION 601235904F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3608089 5',
            mRNA sequence.
ACCESSION   BE515313
VERSION   BE515313.1   GI:9722528
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1268)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM260 row: g column: 02
            High quality sequence start: 35
            High quality sequence stop: 260.
FEATURES   Location/Qualifiers
     source   1..1268
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3608089"
                     /clone_lib="NIH_MGC_44"
                     /tissue_type="endometrium, adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Library constructed by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT   496 a   352 c   332 g   88 t
ORIGIN

Query Match   8.2%; Score 36.6; DB 136; Length 1268;
Best Local Similarity   49.2%; Pred. No. 1.3;
Matches   96; Conservative   0; Mismatches   99; Indels   0; Gaps   0;
```

RESULT 18
AI938961/c
LOCUS   AI938961   175 bp   mRNA   EST   13-DEC-1999
DEFINITION sc63ell.yl Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1016-1125 5' similar to TR:O22152 O22152 F4L23.30 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AI938961
VERSION   AI938961.1 GI:5677831
KEYWORDS   EST.
SOURCE   soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1 (bases 1 to 175)
AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
TITLE   Public Soybean EST Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Trace considered overall poor quality
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 1.
FEATURES   Location/Qualifiers
     source   1..175
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-1125"
                     /clone_lib="Gm-c1016"
                     /tissue_type="immature flowers of field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from immature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT   54 a   45 c   43 g   33 t
ORIGIN

Query Match   7.9%;   Score 35; DB 103; Length 175;
Best Local Similarity 57.9%;   Pred. No. 2.4;
Matches   62; Conservative   0; Mismatches   45; Indels   0; Gaps   0;

Qy 269 ctccggctggagccatttgagtgctgtttaccgtgcagcactcgaagacttccttgctgt 328
        ||| || |||| | ||||| ||| |||| | |||| | |||||| | ||||||| |
Db 107 CTTTCTGTGACTGACTGAATCAGGATTGTTGGCTTGATCCTCGCATCCTCTTCCTTTATGG 48

Qy 329 actcgctggcagcagtagtctgtttgctgcctggtgtctcgggacacctg 375

Db   47 ACTGGTTATGAGCATAAGGTACGCGTGCCTCTTCTCGGGAGGCCTG 1
         ||| | |||||| ||||| | ||||||||| ||| |||| ||||

RESULT 19
AI881530
LOCUS   AI881530   584 bp   mRNA   EST   02-FEB-2000
DEFINITION 606070D01.yl 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI881530
VERSION   AI881530.1 GI:5555579
KEYWORDS   EST.
SOURCE   Zea mays.
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade: Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 584)
AUTHORS   Walbot,V.
TITLE   Maize ESTs from various cDNA libraries sequenced at Stanford
            University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606070 row: D column: 01.
FEATURES   Location/Qualifiers
     source   1..584
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT   143 a   163 c   181 g   97 t
ORIGIN

Query Match   7.9%;   Score 35; DB 103; Length 584;
Best Local Similarity 69.2%;   Pred. No. 3.3;
Matches   63; Conservative   0; Mismatches   25; Indels   3; Gaps   1;

Qy   20 tatgcaggaggccctagtgggt---catctgaaggaaaggacaaagacacctccaagccg 76
        | ||||| |||| |||||| ||| | ||| | ||||| | ||||||| ||| |||||||
Db 482 TTTGCAGAAGCGTCTGGTGGATCAGCAGTCGGGAAAACGGCAAAGACAGCTCGAAGCCG 541

Qy   77 gacacggctactgtagctggcacggcataat 107
        ||| ||| ||||||||||| ||| ||| | ||
Db 542 GACCGCAGCCGCTGGTACCGGCACAGCTTGAT 572

RESULT 20
CNS0062H
LOCUS   CNS0062H   968 bp   DNA   GSS   03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR13D18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061700
VERSION   AL061700.1 GI:4943900
KEYWORDS   GSS.
SOURCE   fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

```
REFERENCE
  AUTHORS    Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  TITLE      Genoscope.
  JOURNAL    Direct Submission
             Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
  COMMENT    Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see http://www.fruitfly.org The BDGP Drosophila
             melanogaster BAC library was prepared by Kazutoyo Osoegawa and
             Aaron Mammoser in Pieter de Jong's laboratory in the Department of
             Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
             NY. The library is named RPCI-98 and was constructed by partial
             EcoRI digestion of Drosophila DNA provided by the BDGP from the
             isogenic strain y2; cn bw sp, the same strain used for the BDGP's
             P1 and EST libraries. A more detailed description of the library
             and how to order individual BAC clones, the entire library, or
             filters for hybridization from the BACPAC Resource Center can be
             found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES            Location/Qualifiers
  source     1..968
             /organism="Drosophila melanogaster"
             /db_xref="taxon:7227"
             /clone_lib="RPCI-98"
             /clone="BACR13D18"
             /note="end : TET3"
BASE COUNT    229 a    149 c    327 g    136 t    127 others
ORIGIN

Query Match
Best Local Similarity   7.7%;   37.8%;   Pred. No. 5.9;   Score 34.4;  DB 219;  Length 968;
Matches   76;  Conservative 25;  Mismatches 100;  Indels 0;  Gaps 0;

Qy  232  tgtttagtcgtagcgtgatgacttcctttgctgtactccgcttggagtg  291
         |||| ||  ||| || |||||| ||||||||||||  ||
Db  115  TTTTTTCTGGTCACGGGCCAGTGGAGTCGGAGTCGGCAATAGAAGCAKAYGAGNGNCCAATGNGTNT  174

Qy  292  ctgtttaccgtgcagcactcgaagacttcctgctgtactccgctggcagcatagtctgtt  351
         | ||||| |  |:: || ||  :|| ||:||:|| ||||||  |
Db  175  CCANTTAGGTGCCACNTNTTNTTWTTKWTAAKDKTKTTCCTAKAWCGCKGSKTKGTCKTTT  234

Qy  352  tgctgcctggtcggggacacctgtactacagtactccgtttccctttgtagtggtac  411
         |:||||:: |  ||| : |||||| ||:: |||||:: ::|:|:: |:|| |:|||:||
Db  235  TGTTTTCTCTTWTAGCGTGCTCTGKRATKCKTKATKMKTKCTTTKTMTTKTATGGKTKGSA  294

Qy  412  tacttcttagccatgtactccg  432
         |:|| ::|| ||| |:: ||
Db  295  TGSTTTGTCGGTGGTKBTTTTG  315


RESULT 21
CNS04AQV/c
LOCUS       CNS04AQV      829 bp    DNA       GSS       21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            096C15 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL282208
VERSION     AL282208.1  GI:8020548
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1 (bases 1 to 829)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizanes,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
```

```
  TITLE      Charaterization and repeat analysis of the compact genome of the
             freshwater pufferfish Tetraodon nigroviridis
  JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 829)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE      Human gene number estimate provided by genome wide analysis using
             Tetraodon nigroviridis DNA sequence
  JOURNAL    Unpublished
REFERENCE   3 (bases 1 to 829)
  AUTHORS   Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
  COMMENT    This sequence is a single read and was generated as part of a large
             scale clone-end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/Tetraodon.
FEATURES            Location/Qualifiers
  source     1..829
             /organism="Tetraodon nigroviridis"
             /db_xref="taxon:99883"
             /clone="096C15"
             /clone_lib="G"
             /note="Genoscope sequence ID : C0BG096AB08SP1-end :
             PUC-Ori"
BASE COUNT    240 a    186 c    174 g    226 t    3 others
ORIGIN

Query Match
Best Local Similarity   7.7%;  56.8%;  Pred. No. 6.5;   Score 34.2;  DB 221;  Length 829;
Matches   63;  Conservative 0;  Mismatches 48;  Indels 0;  Gaps 0;

Qy  221  cattcgtgtgttactgtgtagcgttgatgacttcctgctgtactccgcttggag  280
         || |||| ||| ||| |||::||| || || ||:|| ::|  | |||| ||| |||
Db  556  CAGTCGTTTCAGTATTGTCTTTAGTGGTTGATGTTTCTTNGGCGCCCCAGCTGGGAG  497

Qy  281  ccatttgagtgctgttacggtgcagcactcgaagacttcctgctgtact  331
         ||  |||| ||| ||| |||| |||  ||  :|||:|  |::|| ||| ||
Db  496  ACAGGGCAGGACAGTGTTCTGTGCTGCTGATGTTGACCAGGCACTCGGCTGGACT  446


RESULT 22
AW262166/c
LOCUS       AW262166      364 bp    mRNA      EST       28-DEC-1999
DEFINITION  xq62b06.x1 NCI_CGAP_HN9 Homo sapiens cDNA clone IMAGE:2755187 3'
            similar to contains element LTR5 repetitive element ;, mRNA
            sequence.
ACCESSION   AW262166
VERSION     AW262166.1  GI:6638982
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 364)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index.
  JOURNAL    Unpublished (1997)
  COMMENT    Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Edward Shillitoe Ph.D., Silvio Gutkind Ph.D.,
             Chidchanok Leethanakul D.D.S., Michael Emmert-Buck M.D. Ph.D.
             cDNA Library Preparation: David B. Krizman, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html

             Possible reversed clone: polyT not found
```

```
FEATURES        Seq primer: -40UP from Gibco.
    source      Location/Qualifiers
                1. .364
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:2755187"
                /clone_lib="NCI_CGAP_HN9"
                /tissue_type="normal squamous epithelium from retromolar
                trigone"
                /lab_host="DH10B"
                /note="Vector: pAMP10; cDNA made by oligo-dT priming.
                Non-directionally cloned into the UDG sites of pAMP10.
                Size-selected on agarose gel, average insert  size 500 bp.
                Primary library; non-amplified. cDNA library
                Preparation: David B. Krizman, Ph.D (NCI). Reference:
                Krizman et al. (1996) Cancer Research 56:5380-5383."

BASE COUNT    150 a    134 c    71 g    9 t
ORIGIN

Query Match                 7.7%;    Score 34;   DB 113;   Length 364;
Best Local Similarity     48.9%;    Pred. No. 6;
Matches   91;   Conservative    0;   Mismatches   95;   Indels    0;   Gaps    0;

Qy    202  ggtcgatgttatcggagccattcgtgctgtcttactgtgtagcgcttgatgacttcct  261
           ||| ||   |||| || ||||||||| ||||||  || |||   ||   |||   ||
Db    221  GCTGTGCTGTCTGCAGTCTGCTGCTGTGTTCGTGTCGCGCTGCGTGCTGCTG  162

Qy    262  tgtgctactccgcttggagccatttgagtgtgctgtttaccgtgcagcactcgaagacttcc  321
           |||||| ||   ||| || ||   ||   || |||| |  | ||   || ||
Db    161  TGCTGTGCTGTGCTGCCTGCCGTGCTGTGATGTCTGTGATGTGGTGTTGTCTGTGCTGTTC  102

Qy    322  ttgctgactcgctgcgcagcatagtctgtttgctgcctggtctggggacacctgtact  381
           | |||| ||                   | | |   |||||             ||| ||
Db    101  TGTCTGTGCTGTGCTGTGCTGCTGGTGGCTGCTGCTGTCTGTGCTGTGCTGCGTGTGCTCT  42

Qy    382  acagta 387
           |||
Db    41   GCTGTA 36

RESULT 23
AZ632025
LOCUS       AZ632025        674 bp    DNA               GSS       13-DEC-2000
DEFINITION  1M046A04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0486A04 R, DNA sequence.
ACCESSION   AZ632025
VERSION     AZ632025.1   GI:11754215
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 863)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose.R., Stokes.R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000).
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000    Std Error: 0.00
            Plate: 0486  row: A  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
```

```
FEATURES        High quality sequence stop: 674.
    source      Location/Qualifiers
                1. .674
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0486A04"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: PWD42nv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained  from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a
                0.005 inch orifice at constant velocity. The sheared DNA
                was blunt end-repaired with T4 DNA polymerase and T4
                polynucleotide kinase. Adaptor oligonucleotides were
                ligated to the blunt ends in high molar excess. The
                adaptored DNA was purified and size-selected for a 9.5 to
                10.5 kb range using preparative agarose gel
                electrophoresis. Vector DNA was prepared from a derivative
                of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                inducible derivative of plasmid R1. The vector was ligated
                with adaptors complementary to the insert adaptors and
                purified. The sheared, adaptored mouse DNA was annealed to
                adaptored vector DNA, and transformed into
                chemically-competent E. coli XL10-Gold (Stratagene) cells
                and selected for ampicillin resistance."

BASE COUNT    159 a    127 c    129 g    259 t
ORIGIN

Query Match                 7.7%;    Score 34;   DB 247;   Length 674;
Best Local Similarity     53.8%;    Pred. No. 7.1;
Matches   70;   Conservative    0;   Mismatches   60;   Indels    0;   Gaps    0;

Qy    181  acaagactaacaacgccaatggcgtcgatgttatcggagccattcgtgctgtttactg  240
           ||||  || |  | ||| || |   ||||| |||  |  | ||||| || ||||||
Db    194  ACAAACCTTCCAAGACTCTGCCTATTTATATTCCAGTCATCAATGTGACATTTACAG  253

Qy    241  tgtagcgcttgatgacttccttgctgtactccgcttggagccattgagtgtgctgtttacc  300
           ||| | |||  | ||||| |  |||| |   |||| || ||||  || ||| || | |
Db    254  TGTTTTCCTTAACAAGTGCTTGTTGTTTTCTGCTTTTTGTTATTTCAGCAATGGTAGTG  313

Qy    301  gtgcagcact 310
           || || |  |
Db    314  GTTGAGAAAT 323
```

```
RESULT 24
CNS013VO
LOCUS       CNS013VO        863 bp    DNA               GSS       26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN10C13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103374
VERSION     AL103374.1   GI:5614985
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 863)
AUTHORS     Direct Submission
TITLE       Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL     BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
```

```
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11

FEATURES             Location/Qualifiers
     source          1..863
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN10C13"
                     /note="end : SP6"
BASE COUNT      86 a     167 c     186 g     259 t     165 others
ORIGIN

Query Match        7.6%;   Score 33.8;  DB 219;  Length 863;
Best Local Similarity  35.6%;  Pred. No. 8.9;
Matches   78;  Conservative  33;  Mismatches 108;  Indels   0;  Gaps   0;

Qy  109 gaaaacagtccctgcccttttgagggtcgtctcaaacatgcccctcgagacgtcggtgaa 168
        || ||| ||||     ||| |||| : ||||| || || : |||||       | |  ||
Db  564 GANTGCAGNCTATMTTTGATGGMTGGTTGTTAATAGTATTTGCTTCGTTMKTTTTTTKGMK 623

Qy  169 gccgtagttgtacaagactaacacgccaatggcgtcgatgtatcggagccattcgtg 228
        :||  |   ||    |  : | ||  || |  |||  ||||||||| |    ||  :|
Db  624 GATGTKMTTTMTTTKKCMMTGKMTMAMTTGTKGGGGTCCMTGTGCGMTGMMGKTTCMMMTK 683

Qy  229 tgcgtgttactgtgtagcgcttgatgacttccttgctgtactccgctggagccatttga 288
        :|: : ||:  | || | |   || ||||||||||| |: ||| :||  |  |||||
Db  684 MTMTTTGKGMTGKTKKGTCGKTGTTTTNMTTTTTKCCGTTMTTTTTCYGTTTTCTTTTG 743

Qy  289 gtgctgttaccgtgcagcactcgaagacctccttgctg 327
        :|:| || || |: ||| :|| ||| |||||||||| ||
Db  744 CTKCYCMTTTCMGGGTTCMGCTTSWGBMMTKYGTGGGTG 782

RESULT 25
CNS00FEV
LOCUS       CNS00FEV      1101 bp    DNA        GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR31J08 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL070299
VERSION     AL070299.1   GI:4950527
KEYWORDS    GSS.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp; the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
```

```
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11

FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR31J08"
                     /note="end : T7"
BASE COUNT     260 a     200 c     206 g     217 t     218 others
ORIGIN

Query Match        7.6%;   Score 33.6;  DB 219;  Length 1101;
Best Local Similarity  26.5%;  Pred. No. 11;
Matches   54;  Conservative  61;  Mismatches  89;  Indels   0;  Gaps   0;

Qy  189 aacaacgccaatggcgtcgatgtatcggagccattcgtgtcgtgttactgtgtagcgc 248
        ||::  ||  : || |: || | :  :|| ||: |::|:  : || |||  ||  |  ::
Db  896 AAMHGKTYCCRKGGYKKGBDYGCGTAYASARVCCTGGATAGYGTTTCAAARSWATASCSY 955

Qy  249 ttgatgacttccttgctgtactccgcttggagccatttgagtgctgttaccgtgcagca 308
        ||| |:|: ||  | :||||:|  |  ::|||| ||| |:  |: |: || ||:|| |||
Db  956 TTKTTGMTTHWVGSGSGCKADAWCGCATTKMDTKTYBCTAATAGCCCACACTGARGGGCT 1015

Qy  309 ctcgaagacttcctgctgtactgctggcagcatagtcgtggcgtgctggtcgggg 368
        :| |||| || ::| |||||:| | :|:|| || |: |:| || |||| ||| :|
Db  1016 GRGSGACAKNNSMSYGBKRTAMCCTGGGGCGMGCWCDYGGSASCGTTYGYCSBGGCVSC 1075

Qy  369 acacctgtactacagtacttcg 392
        :|:| : | ||| || ||:| :|
Db  1076 MSAYGYGYGYKCTCCTKGWSAKCG 1099

RESULT 26
AI685592/c
LOCUS       AI685592                        EST       15-DEC-1999
DEFINITION  tu20c07.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2251596 3'
            similar to contains PTR5.t3 MSR1 repetitive element :, mRNA
            sequence.
ACCESSION   AI685592
VERSION     AI685592.1   GI:4896886
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 386)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1143   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 380.
FEATURES             Location/Qualifiers
     source          1..386
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2251596"
                     /clone_lib="NCI_CGAP_Pr28"
                     /sex="male"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: prostate; Vector: pT7T3D-Pac (Pharmacia)
```

with a modified polylinker; Plasmid DNA from the
normalized library NCI_CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneIDs
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."

```
BASE COUNT      72 a    140 c    110 g     64 t
ORIGIN
```

```
Query Match
Best Local Similarity    7.5%;    Score 33.4;    DB 23;    Length 386;
Matches    58;    Conservative    0;    Mismatches    41;    Indels    0;    Gaps    0;

Qy    290 tgctgtttaccgtgcagcactgcgaagacttcctgtgtactcgctggcagcatagtctg 349
          ||| ||| ||  |||      | ||||| |||||||  |||| | |||||||||||||
Db    188 TGCTCAGACCCGGCTGGCGCTGCCCATGCGCTGCGTTTGGCGTGGCCTGGCCTGTGTCTG 129

Qy    350 tttgctgcctggtctggggacacctgtgtactacagtac 388
          || ||||||||||| | |    || |  | |||| |
Db    128 TTAGCTGCCTGGTGCCCAGGGAGCTGTCTGCTGCAGTGC 90
```

```
RESULT 27
BEB21365
LOCUS       BE821365    412 bp    mRNA    EST    21-SEP-2000
DEFINITION  GM700024A20H8 Gm-r1070 Glycine max cDNA clone Gm-r1070-3808 3',
            mRNA sequence.
ACCESSION   BE821365
VERSION     BE821365.1    GI:10253599
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1 (bases 1 to 412)
AUTHORS     Vodkin,L.,Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
TITLE       A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL     Unpublished (1999)
COMMENT     Other_ESTs: AI938961 corresponding to Gm-c1016-1125 (5')
            Contact: Vodkin,L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES    Location/Qualifiers
     source 1..412
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="Gm-r1070-3808"
            /clone_lib="Gm-r1070"
            /note="The library Gm-r1070 is a sequence-driven, reracked
            set of 9,216 clones selected from cDNA libraries from
            various tissues and stages of development of soybean that
            represent 2,639 sequences from immature cotyledons, 1,770
            from immature seed coats, 3,938 from flowers, and 869
            from young pods. The 5' ESTs of the source clones from
            the different libraries was used to select singletons, or
            a representative of each contig, which were reracked to
```

form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Keck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

```
BASE COUNT     112 a     88 c     82 g    125 t      5 others
ORIGIN
```

```
Query Match
Best Local Similarity    7.5%;    Score 33.4;    DB 140;    Length 412;
Matches    64;    Conservative    0;    Mismatches    51;    Indels    0;    Gaps    0;

Qy    269 ctccgcttggagccattgagtgctgttaccgtgcagcactcgaagacttcctgctgt 328
          ||| ||  |    |||   |||| | ||   | | ||  ||||||||| ||| ||
Db    265 CTTCTCTGTGACTGATATCAGGATTATTGGCTTGATCCTCTGAATCCTCTTCCTTTATAA 324

Qy    329 actcgctggcagcatagtctgtttgctgcctggtctggggacacctgtgtactac 383
          ||| |  | | ||  ||||| |     ||| |||| ||| |||  |  | |
Db    325 ACTGTTTATAAGCAGAAGTACGGCGTGCCTCTTCTCGGGAGGCCTGTTCACAAC 379
```

```
RESULT 28
BG508825/c
LOCUS       BG508825    523 bp    mRNA    EST    28-MAR-2001
DEFINITION  sac90g12.y1 Gm-c1073 Glycine max cDNA clone GENOME SYSTEMS CLONE
            ID: Gm-c1073-336 5' similar to TR:Q9XFB1 Q9XFB1 YABBY3. ; mRNA
            sequence.
ACCESSION   BG508825
VERSION     BG508825.1    GI:13479482
KEYWORDS    EST.
SOURCE      soybean.
ORGANISM    Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1 (bases 1 to 523)
AUTHORS     Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
TITLE       Public Soybean EST Project
JOURNAL     Public Soybean EST Project
            Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 383.
FEATURES    Location/Qualifiers
     source 1..523
            /organism="Glycine max"
```

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1810
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 419.

FEATURES   Location/Qualifiers
source   1..535
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1069-621"
                     /clone_lib="Gm-c1069"
                     /tissue_type="Degenerating cotyledons, 9-10 day old
                     etiolated seedling"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from degenerating cotyledons of 9-10 day old etiolated
                     seedlings for the cultivar Williams. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with a XhoI restriction site. EcoRI
                     adapters were ligated to the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA fragments were
                     directionally cloned into the EcoRI-XhoI restriction site
                     of the pBluescript vector. The ligated cDNA fragments
                     were transformed into DH10B host cells (GibcoBRL). This
                     library was constructed in the laboratory of Dr. Randy
                     Shoemaker."

BASE COUNT   139 a   128 c   133 g   134 t   1 others
ORIGIN

Query Match   7.5%;   Score 33.4;   DB 148;   Length 535;
Best Local Similarity 55.7%; Pred. No. 10;
Matches   64; Conservative   0; Mismatches   51; Indels   0; Gaps   0;

Qy   269 ctccgcttggagccattgagtgctgttaccgtgcagcactcgaagacttccttgctgt 328
         |||| || |||| |||||||||| ||| || || || |||||| |||||||| ||||||
Db   460 CTTCCTGTGACTGATATCAGGATTATTGGCTTTGATCCTCTGAATCTCTTCCTTATAA 401

Qy   329 actgctggcagcatagtctgtttgctgcctggtctggggacccctgtgtactac 383
         ||| ||||| | || ||||| |||||||| ||| || | |||| ||| |||||
Db   400 ACTGGTTATAAGCAGAAGGTACGCGCTGCCTCTTCGGGAGGCCTGTTCACAAC 346

RESULT 30
AW733988/c
LOCUS   AW733988   544 bp   mRNA   EST   24-APR-2000
DEFINITION sk79b06.y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1016-10284 5' similar to TR:O22152 O22152 F4L23.30 PROTEIN. ;,
            mRNA sequence.
ACCESSION   AW733988
VERSION   AW733988.1 GI:7639670
KEYWORDS   EST.
SOURCE   soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1 (bases 1 to 544)
AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
TITLE   Public Soybean EST Project
JOURNAL   Unpublished (1999)

-- (left column) --

                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1073-336"
                     /clone_lib="Gm-c1073"
                     /tissue_type="seedlings induced for symptoms of SDS
                     (Sudden Death Syndrome) disease"
                     /dev_stage="2-3 weeks old"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from 2-3 week old seedlings that were induced for symptoms
                     of SDS (Sudden Death Syndrome) disease by the
                     translocation of culture filtrate of Fusarium solani f.
                     sp. glycines (Plant Cell Report 18:375-380). Cultivar
                     Williams 82 is susceptible to the disease SDS. Plant
                     tissue (expanded leaves, folded leaves, and new shoots)
                     were collected at 1, 6, 24, and 48 hrs. after inoculation
                     and their mRNA pooled equally for cDNA construction. The
                     library was prepared using the Stratagene pBluescript II
                     SK(+) library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with an XhoI restriction site. EcoRI
                     adaptors were ligated to the blunt-ended cDNA fragments
                     followed by XhoI digestion. The cDNA insert is protected
                     from XhoI digestion via methylation during first strand
                     synthesis. The cDNA fragments were directionally cloned
                     into the EcoRI-XhoI restriction site of the pBluescript
                     vector. The ligated cDNA fragments were transformed into
                     E.coli ElectroMax DH10B host cells. Plants were inoculated
                     by Shuxian Li (Glen Hartman lab, University of Illinois).
                     Library was constructed by Reena Philip and Steve Clough
                     (Lila Vodkin lab, University of Illinois)."

BASE COUNT   143 a   120 c   131 g   129 t
ORIGIN

Query Match   7.5%;   Score 33.4;   DB 154;   Length 523;
Best Local Similarity 55.7%; Pred. No. 10;
Matches   64; Conservative   0; Mismatches   51; Indels   0; Gaps   0;

Qy   269 ctccgcttggagccattgagtgctgttaccgtgcagcactcgaagacttccttgctgt 328
         |||| || |||| |||||||||| ||| || || || |||||| |||||||| ||||||
Db   404 CTTCCTATGACTGATATCAGGATTATTGGCTTTGATCCTCTGAATCTCTTCCTTATAA 345

Qy   329 actgctggcagcatagtctgtttgctgcctggtctggggacccctgtgtactac 383
         ||| ||||| | || ||||| |||||||| ||| || | |||| ||| |||||
Db   344 ACTGGTTATAAGCAGAAGGTACGCGCTGCCTCTTCGGGAGGCCTGTTCACAAC 290

RESULT 29
BF424911/c
LOCUS   BF424911   535 bp   mRNA   EST   28-NOV-2000
DEFINITION su52ell.y1 Gm-c1069 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1069-621 5' similar to TR:Q9XFB1 Q9XFB1 YABBY3. ;, mRNA
            sequence.
ACCESSION   BF424911
VERSION   BF424911.1 GI:11412900
KEYWORDS   EST.
SOURCE   soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1 (bases 1 to 535)
AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
TITLE   Public Soybean EST Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project

COMMENT   Contact: Shoemaker R/Public Soybean EST Project
             Public Soybean EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             This clone is available through: Genome Systems, Inc. 4633 World
             Parkway Circle St. Louis, Missouri 63134 For further information
             call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
             427-3324 or contact: clones@genomesystems.com or
             info@genomesystems.com web site: www.genomesystems.com
             Seq primer: -40RP from Gibco
             High quality sequence stop: 422.
FEATURES   Location/Qualifiers
    source   1..544
             /organism="Glycine max"
             /db_xref="taxon:3847"
             /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-10284"
             /clone_lib="Gm-c1016"
             /lab_host="XL10-Gold"
             /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
             XhoI; This cDNA library was constructed from mRNA isolated
             from immature flowers of field grown plants. The cDNA
             library was prepared using the Stratagene pBluescript II
             XR library construction kit. Complementary DNA was
             synthesized from mRNA using a primer consisting of a poly
             (dT) sequence with a XhoI restriction site. EcoRI adapters
             were ligated to the blunt-ended cDNA fragments followed by
             XhoI digestion. The cDNA fragments were directionally
             cloned into the EcoRI-XhoI restriction site of the
             pBluescript vector. The ligated cDNA fragments were
             transformed into XL10-Gold host cells. This library was
             constructed by Dr. Randy Shoemaker and Dr. John
             Erpelding."
BASE COUNT   161 a   106 c   112 g   165 t
ORIGIN

Query Match   7.5%;   Score 33.4;   DB 120;   Length 544;
Best Local Similarity   55.7%;   Pred. No. 10;
Matches   64;   Conservative   0;   Mismatches   51;   Indels   0;   Gaps   0;

Qy   269 ctccgcttgagccattgagtgctgtttaccgtgcagcactcgaagacttccttgctgt 328
           || |-|| ||||- ||-|||| || ||||-|-|| ||-|| |||||||| |
Db   152 CTTCTCTATGACTGATATCAGGATTATTGGCTTGATCCTCTGAATCTCTTCCTTTATAA 93

Qy   329 actgcgctggcagcatagtctgtttgctgcctggctgtctgggacaccctgtactac 383
           ||||| | ||||| | ||||||| | | |||||-||| ||-||| || ||
Db   92   ACTGGTTATAAGCAGAAGTACGCGCTGCCTCCTTCTCGGAGGCCTGGTTCACAAC 38

RESULT 31
AZ380524/c
LOCUS   AZ380524   628 bp   DNA   GSS   02-OCT-2000
DEFINITION   1M013GD02R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
             clone UUGC1M0136D02 R, DNA sequence.
ACCESSION   AZ380524
VERSION   AZ380524.1   GI:10494224
KEYWORDS   GSS.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 628)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
             Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
             ,M., Rose.M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
             and Wright,D.,Weiss,R.
TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
             plasmid inserts
JOURNAL   Unpublished (2000)

COMMENT   Contact: Robert B. Weiss
             University of Utah Genome Center
             University of Utah
             Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
             84112, USA
             Tel: 801 585 5606
             Fax: 801 585 7177
             Email: ddunn@genetics.utah.edu
             Insert Length: 10000   Std Error: 0.00
             Plate: 0136   row: D column: 02
             Seq primer: CACACAGGAAACAGCTATGACC
             Class: plasmid ends
             High quality sequence stop: 628.
FEATURES   Location/Qualifiers
    source   1..628
             /organism="Mus musculus"
             /strain="C57BL/6J"
             /db_xref="taxon:10090"
             /clone="UUGC1M0136D02"
             /clone_lib="Mouse 10kb plasmid UUGC1M library"
             /sex="Male"
             /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
             /note="Vector: PWD42nv; Purified genomic DNA from M.
             musculus C57BL/6J (male) was obtained from the Jackson
             Laboratory Mouse DNA Resource
             (http://www.jax.org/resources/documents/dnares/). The DNA
             was hydrodynamically sheared by repeated passage through a
             0.005 inch orifice at constant velocity. The sheared DNA
             was blunt end-repaired with T4 DNA polymerase and T4
             polynucleotide kinase. Adaptor oligonucleotides were
             ligated to the blunt ends in high molar excess. The
             adaptored DNA was purified and size-selected for a 9.5 to
             10.5 kb range using preparative agarose gel
             electrophoresis. Vector DNA was prepared from a derivative
             of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
             inducible derivative of plasmid R1. The vector was ligated
             with adaptors complementary to the insert adaptors and
             purified. The sheared, adaptored mouse DNA was annealed to
             adaptored vector DNA, and transformed into
             chemically-competent E. coli XL10-Gold (Stratagene) cells
             and selected for ampicillin resistance."
BASE COUNT   166 a   177 c   172 g   113 t
ORIGIN

Query Match   7.5%;   Score 33.4;   DB 242;   Length 628;
Best Local Similarity   52.5%;   Pred. No. 11;
Matches   73;   Conservative   0;   Mismatches   66;   Indels   0;   Gaps   0;

Qy   282 catttgagtgctgtttaccgtgcagcactcgaagacttccttgctgtactcgctggcagc 341
           ||-|| || || |-||||| |||| |-|||| ||| || |||||||| | |
Db   579 CAGTGATGTGCGGCTTGTTATTCTGCTTGCGGAGGATGCCGTCTCTGTACCGCTCTGCAGC 520

Qy   342 atagtctgtttgctgcctgctgtctgggacacctgtgtactacagtacttcgtttccttt 401
           ||||| | |||||||||| |||||||||| | | ||| ||-||| || ||-||-|| ||
Db   519 TGTGTCTCTCACAACCATTGCTTGTCAAGACTTCAGACAGACGGTACTTGGTCTCCCATT 460

Qy   402 gtagtggtactacttctta 420
           | ||| ||| ||| ||||
Db   459 GATGTGCTCTACCTACTAA 441

RESULT 32
BE563288/c
LOCUS   BE563288   933 bp   mRNA   EST   15-AUG-2000
DEFINITION   601135410F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3689402 5',
             mRNA sequence.
ACCESSION   BE563288
VERSION   BE563288.1   GI:9807100
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 933)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM384 row: c column: 03
High quality sequence stop: 678.

FEATURES
source

Location/Qualifiers
1..933
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3689402"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."

BASE COUNT
ORIGIN

148 a   316 c   319 g   150 t

Query Match   7.5%;   Score 33.4; DB 137; Length 933;
Best Local Similarity 58.6%;   Pred. No. 12;
Matches   58; Conservative   0; Mismatches   41; Indels   0; Gaps   0;

Qy   290 tgctgttacgtgcagcactcgaagacttcctgctgtactcgctggcagcatagtctg 349
           ||| ||| ||| || | ||||| || | |||| ||| |||| || |||||||
Db   422 TGCTCAGACCCGGGCTGGCGCTGGCGCTGCCATGCCGCTCGTGTTGGGCGGCTGTGTCTG 363

Qy   350 tttgctgcctggtctggggacacctgtgtactacagtac 388
           || |||||||||| | |||| || |||| |||||||
Db   362 TTAGCTGCCTGGTGTCCCAGGGAGCTGTCTGCTGCAGTGC 324

RESULT 33
BE871340/c
LOCUS   BE871340   945 bp   mRNA   EST   20-OCT-2000
DEFINITION 601449457F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3853334 5',
            mRNA sequence.
ACCESSION   BE871340
VERSION   BE871340.1 GI:10320116
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 945)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:

---

http://image.llnl.gov
Plate: LLAM9577 row: a column: 15
High quality sequence stop: 656.

FEATURES
source

Location/Qualifiers
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3853334"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT
ORIGIN

170 a   363 c   300 g   112 t

Query Match   7.5%;   Score 33.4; DB 141; Length 945;
Best Local Similarity 58.6%;   Pred. No. 12;
Matches   58; Conservative   0; Mismatches   41; Indels   0; Gaps   0;

Qy   290 tgctgttacgtgcagcactcgaagacttcctgctgtactcgctggcagcatagtctg 349
           ||| ||| ||| || | ||||| || | |||| ||| |||| || |||||||
Db   392 TGCTCAGACCCGGGCTGGCGCTGGCGCTGCCATGCCGCTCGTGTTGGGCGGCTGTGTCTG 333

Qy   350 tttgctgcctggtctggggacacctgtgtactacagtac 388
           || |||||||||| | |||| || |||| |||||||
Db   332 TTAGCTGCCTGGTGTCCCAGGGAGCTGTCTGCTGCAGTGC 294

RESULT 34
AV327005/c
LOCUS   AV327005   225 bp   mRNA   EST   11-NOV-1999
DEFINITION AV327005 RIKEN full-length enriched, adult male medulla oblongata
            Mus musculus cDNA clone 6330416I19 3', mRNA sequence.
ACCESSION   AV327005
VERSION   AV327005.1 GI:6367057
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 225)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T.,
            Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F.,
            Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Izawa,M., Kagawa,I., Kai
            ,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M.,
            Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
            Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata
            ,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H.,
            Suzuki,H., Takahashi,F., Tateno,M., Tominaga,N., Tsunoda,Y.,
            Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T.,
            Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al. 1999)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Yoshihide Hayashizaki
            Genome Exploration Research Group, Life Science Tsukuba Center,
            Genome Science Laboratory
            The Institute of Physical and Chemical Research (RIKEN), Genomic
            Sciences Center
            3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
            Tel: +81-298-36-9013
            Fax: +81-298-36-9098
            Email: genome-res@rtc.riken.go.jp,
            URL:http://genome.rtc.riken.go.jp/,
            Sasaki,N., Izawa,M., Watahiki,M., Ozawa,K., Tanaka,T., Yoneda,Y.,
            Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki
            ,Y.
            'Transcriptional sequencing: A method for DNA sequencing using RNA
            polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,

COMMENT   Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES   Location/Qualifiers
source   1..225
     /organism="Mus musculus"
     /strain="C57BL/6J"
     /db_xref="taxon:10090"
     /clone="6330416I19"
     /clone_lib="RIKEN full-length enriched, adult male medulla
oblongata"
     /sex="male"
     /tissue_type="medulla oblongata"
     /dev_stage="adult"
     /lab_host="DH10B"
     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence (5' GAGAGAGAGATTCGAGTTAATTAAATTAATCCCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
BamHI"

BASE COUNT   97 a   33 c   40 g   55 t
ORIGIN

Query Match   7.5%;   Score 33.2;   DB 29;   Length 225;
Best Local Similarity   57.8%;   Pred. No. 9.4;
Matches   59;   Conservative   0;   Mismatches   43;   Indels   0;   Gaps   0;

Qy   189   aacaacgccaatggcgtcgatgttatcggagccattcgtgtgtttactgttgtagcgc 248
     ||||| | |||| |||| |||| | | || | |||||||| || ||| |||
Db   105   AACAATTACCTTGGCTTCGAATTCCCCTGGTTCCCTCGTGTTCTTCATGGTTTATGAC 46

Qy   249   ttgatgacttccttgtgtactccgcttggagccatttgagt 290
     ||||| ||| ||||| | | |||||| ||
Db   45   TTGATATCTTTTGTTGCTTTCTGCAATTTCGCGCGATTAATGT 4

RESULT 35
AI646741
LOCUS   AI646741   287 bp   mRNA   EST   29-APR-1999
DEFINITION   ub65c02.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1382594 3', mRNA sequence.
ACCESSION   AI646741
VERSION   AI646741.1   GI:4725216
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 287)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)

COMMENT   Contact: Robert.Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:905062
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 286.

FEATURES   Location/Qualifiers
source   1..287
     /organism="Mus musculus"
     /db_xref="taxon:10090"
     /clone="IMAGE:1382594"
     /clone_lib="Soares_mammary_gland_NMLMG"
     /sex="female (lactating)"
     /tissue_type="mammary gland"
     /lab_host="DH10B"
     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   63 a   58 c   44 g   122 t
ORIGIN

Query Match   7.5%;   Score 33.2;   DB 23;   Length 287;
Best Local Similarity   51.3%;   Pred. No. 10;
Matches   77;   Conservative   0;   Mismatches   73;   Indels   0;   Gaps   0;

Qy   44   tctgaaggaaaggacaaagacacctccaagccggacacggctactgtagctggcacggca 103
     ||| || ||| ||| || || | || | |||||||||||||||||||| |||| | |
Db   103   TTTGGGGGGAAAAAATAAAACCCCCCCCCCCACCCAAGGAATTTTGGTTTGGGGGGG 162

Qy   104   taatggaaaacagtccctgccctttgagggtcgtctcaaacatgcccctcgagacgtcg 163
     ||| | ||| | ||| || ||| |||| |||| || | ||| | ||| |||
Db   163   GGAAAAACCCAAAACCCCCCCCCCTTTGGGGTCCCCCAAAAAACCCCCGGGGGGTTG 222

Qy   164   gtgaagccgtagtttgtacaagactaacca 193
     ||- ||- || || ||||| || |||
Db   223   GGGACCCCCCTTTTTTTTTAAAAAAAACCCA 252

RESULT 36
AZ640083/c
LOCUS   AZ640083   742 bp   DNA   GSS   14-DEC-2000
DEFINITION   1M0502K04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0502K04 F, DNA sequence.
ACCESSION   AZ640083
VERSION   AZ640083.1   GI:11763579
KEYWORDS   GSS.
SOURCE   house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 742)
AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Email: ddunn@genetics.utah.edu
Insert Length: 10000   Std Error: 0.00
Plate: 0502   row: K   column: 04
Seq primer: CGTTGTAAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 742.

```
FEATURES       Location/Qualifiers
source         1..742
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="UUGC1M0502K04"
               /clone_lib="Mouse 10kb plasmid UUGC1M library"
               /sex="Male"
               /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
               /note="Vector: PWD42nv; Purified genomic DNA from M.
               musculus C57BL/6J (male) was obtained from the Jackson
               Laboratory Mouse DNA Resource
               (http://www.jax.org/resources/documents/dnares/). The DNA
               was hydrodynamically sheared by repeated passage through a
               0.005 inch orifice at constant velocity. The sheared DNA
               was blunt end-repaired with T4 DNA polymerase and T4
               polynucleotide kinase. Adaptor oligonucleotides were
               ligated to the blunt ends in high molar excess. The
               adaptored DNA was purified and size-selected for a 9.5 to
               10.5 kb range using preparative agarose gel
               electrophoresis. Vector DNA was prepared from a derivative
               of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
               inducible derivative of plasmid RI. The vector was ligated
               with adaptors complementary to the insert adaptors and
               purified. The sheared, adaptored mouse DNA was annealed to
               adaptored vector DNA, and transformed into
               chemically-competent E. coli XL10-Gold (Stratagene) cells
               and selected for ampicillin resistance."
BASE COUNT     312 a    265 c    153 g    12 t
ORIGIN
```

Query Match   7.5%;   Score 33.2;   DB 247;   Length 742;
Best Local Similarity 48.4%;   Pred. No. 13;
Matches 92;   Conservative 0;   Mismatches 98;   Indels 0;   Gaps 0;

```
Qy   202  ggcgatgatgttatcggagccattcgtgtgctgttcactgtgtagcgcttgatgacttcct  261
          ||  |||  |  ||   ||  | || |||||  ||  ||  |  ||  ||  |  ||  || ||
Db   249  GCTGTGCTGTCGTGCAGTCTGCTCGTGCCGTGTGCGTGTTGCTGCTCGTGCCGCTGCCTG  190

Qy   262  tgctgtactccgcttggaccattgagtgctgttacctgcagcactcgaagacttcc  321
          |||||||| |||  |  |  |  | |  ||  |  |  ||  |  |||  ||  |  |
Db   189  TGCTGTGCTGTCGTGCTCGCCGTGCTCGTGATGTCGTGATGGTGTGTGTCGTGCTGCGTTC  130

Qy   322  ttgctgtactccgctggacagcatagtctgtttgctgcctgctgggacacctgtgtact  381
          | |||| ||  | ||   |  |   |  |  | ||| | ||  |||||||||  || |
Db   129  TGTCTGTGCTGTGCTGTCGTGCTGCGTGCTCGTGCTGCGTGCTGCTGTGCGTCTGTGCTCT  70

Qy   382  acagtacttc  391
          | |||  ||
Db   69   GCTGTATGTC  60
```

RESULT 37
AW146599
LOCUS   AW146599   672 bp   mRNA   EST   03-NOV-1999
DEFINITION 614074C07.y3 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW146599
VERSION   AW146599.1   GI:6194495
KEYWORDS   EST.
SOURCE   Zea mays.
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 672)

AUTHORS   Walbot,V.
TITLE   Maize ESTs from various cDNA libraries sequenced at Stanford
           University
JOURNAL   Unpublished (1999)
COMMENT   Contact: Walbot V
           Department of Biological Sciences
           Stanford University
           855 California Ave, Palo Alto, CA 94304, USA
           Tel: 650 723 2227
           Fax: 650 725 8221
           Email: walbot@stanford.edu
           Plate: 614074 row: C column: 07.

```
FEATURES       Location/Qualifiers
source         1..672
               /organism="Zea mays"
               /cultivar="W23"
               /db_xref="taxon:4577"
               /clone_lib="614 - root cDNA library from Walbot Lab"
               /tissue_type="root"
               /dev_stage="3-4 days old"
               /lab_host="XL1OLR"
               /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
               EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
               Lab (LM)"
BASE COUNT     181 a    139 c    168 g    184 t
ORIGIN
```

Query Match   7.4%;   Score 33;   DB 112;   Length 672;
Best Local Similarity 58.8%;   Pred. No. 15;
Matches 57;   Conservative 0;   Mismatches 40;   Indels 0;   Gaps 0;

```
Qy   319  tccttgctgtactcgctggcagcagtctgtttgctgcctggtctgggaccacctgt  378
          | || |||| |||||  ||  |   |  || | || |||||  ||  |  |||  | || ||
Db   546  TTCTGGATTTATTCCGTAGCACCATAGTTATTTTTGGTTTGACATGGAGGACATTGTG  605

Qy   379  actacagtacttcgtttcccctttgtagtggtactact  415
          ||||||| ||| || |||  ||  ||  |||  | ||| ||
Db   606  ACTACAGAACTGAGTTTACTATCCATAACACTGCT  642
```

RESULT 38
AQ180148/c
LOCUS   AQ180148   416 bp   DNA   GSS   23-OCT-1998
DEFINITION HS_3200_B1_C04_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3200 Col=7 Row=F, DNA sequence.
ACCESSION   AQ180148
VERSION   AQ180148.1   GI:3577515
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 416)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
           Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
           Hood,L.
TITLE   Sequence-tagged connectors: A sequence approach to mapping and
           scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
           High Throughput Sequencing Center
           University of Washington
           401 Queen Anne Avenue North, Seattle, WA 98109, USA
           Tel: (206) 616-3618
           Fax: (206) 616-3887
           Email: jwallace@u.washington.edu
           Sequence Tagged Connector
           Plate: 3200 row: F column: 7
           Class: BAC ends
           High quality sequence stop: 416.
FEATURES   Location/Qualifiers
```

```
source          1..416
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="Plate-3200 Col=7 Row=F"
                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                /sex="male"
                /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                E-Coli DH10B"

BASE COUNT    114 a    86 g    106 c    104 t    6 others
ORIGIN

Query Match         7.4%;    Score 32.8;  DB 225;  Length 416;
Best Local Similarity  57.3%;  Pred. No. 15;
Matches  55;  Conservative  0;  Mismatches  41;  Indels  0;  Gaps  0;

Qy  135  tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
         |||||| ||||| ||  || ||||||||  || ||||||||||| |  ||||||
Db  415  TCCTNTAAAGAAACCNNTGCGAAATGTCTGGGTACACNTGGATGGTCCTGGAGTAACAAG 356

Qy  195  gccaatggcgtcgatgttatcggagccattcgtgtg 230
         ||||| |||||||  ||||| || |||||| |
Db  355  ACCAATGAAGAAGAGATGTTCCGGTAAACCTTGGTTG 320

RESULT 39
BF974202/c
LOCUS       BF974202    715 bp    mRNA    EST    22-JAN-2001
DEFINITION  602243835F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4334961 5',
            mRNA sequence.
ACCESSION   BF974202
VERSION     BF974202.1  GI:12341417
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1205  row: e  column: 10
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1..715
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4334961"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT    153 a    215 g    238 c    109 t
ORIGIN


Query Match         7.4%;    Score 32.8;  DB 172;  Length 715;
Best Local Similarity  54.0%;  Pred. No. 17;
Matches  67;  Conservative  0;  Mismatches  57;  Indels  0;  Gaps  0;

Qy   23  gcaggagcgcctagtgggtcatctgaaggaaggacaaagacacctcaagccggacacg 82
         ||| ||||  ||||| | ||| | |||||||| |||||||||| |||||| ||||||
Db  375  GGAACAGTGCTCAGTTCAGTTCATGGTCTGCAGGATACCTTCTGCACAGCCTCCAGCCCGGCTGG 316

Qy   83  gctactgtagctggcacggcatcatgaaaacagtccctgcccctttgagggtcgtctca 142
         ||||| || || || |||| ||  || |||| |  ||||| | |||||| | ||  |||
Db  315  GGGAACGGTGTTGTCACGGCACACGAGCGGGTAGCCCTCTCGCCCTTGGTCCGGGGATCTTC 256

Qy  143  aaca 146
         ||||
Db  255  AACA 252

RESULT 40
BF575314/c
LOCUS       BF575314    887 bp    mRNA    EST    12-DEC-2000
DEFINITION  602134043F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4288906 5',
            mRNA sequence.
ACCESSION   BF575314
VERSION     BF575314.1  GI:11649026
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 887)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM129  row: f  column: 11
            High quality sequence stop: 556.
FEATURES             Location/Qualifiers
     source          1..887
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4288906"
                     /clone_lib="NIH_MGC_81"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                     (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI
                     (ggccattatggcc); 5' and 3' adaptors were used in cloning
                     as follows: 5' adaptor sequence: 5'-CACGGGCCATTATGGCC-3'
                     and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size
                     1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech Laboratories (Palo
                     Alto, CA)."
FEATURES
     source          1..887
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4288906"
                     /clone_lib="NIH_MGC_81"
BASE COUNT    270 a    195 g    155 c    265 t    2 others
ORIGIN


Query Match         7.4%;    Score 32.8;  DB 150;  Length 887;
Best Local Similarity  54.0%;  Pred. No. 19;
Matches  67;  Conservative  0;  Mismatches  57;  Indels  0;  Gaps  0;

Qy  195  gccaatggcgtcgatgttatcggagccattcgtgctgcttactgtgttagcgcttgatg 254
Db  810  GGCTATGGTGCTGGTGGTTGAGTCCTCAACCGTGTTCTGTTTATAGGCCTATCCTCTGCG 751
```

```
Qy    255  acttccttgctgtactccgcttggagccattgagtgctgtttaccgtgcagcactcgaa  314
           ||  ||||||  |||| || ||||| ||  || ||| ||| ||| || ||| |||| ||||
Db    750  AGGGCCTTGTGTTCCTTCGCCTGCCCCAGTTGATGCCCTCTATAGATCCGTTCTCTAA   691

Qy    315  gact 318
           ||||
Db    690  ACCT 687
```

RESULT 41
BG399947/c
LOCUS   BG399947   1564 bp   mRNA   EST   12-MAR-2001
DEFINITION   602441492F1 NIH_MGC_75 Homo sapiens cDNA clone IMAGE:4557358 5',
              mRNA sequence.
ACCESSION   BG399947
VERSION   BG399947.1 GI:13293395
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1564)
AUTHORS   NIH-MGC http://mgc.ncl.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: CLONTECH Laboratories, Inc.
              cDNA Library Preparation: CLONTECH Laboratories, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLCM1261 row: g column: 23
              High quality sequence stop: 151.
FEATURES   Location/Qualifiers
     source   1..1564
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4557358"
                     /clone_lib="NIH_MGC_75"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.65
                     kb (range 0.5-4.0 kb). 15/15 colonies contained inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT   681 a   444 c   273 g   166 t
ORIGIN

Query Match   7.4%;   Score 32.8; DB 153; Length 1564;
Best Local Similarity 55.2%; Pred. No. 22;
Matches 64; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy   292 ctgttaccgtgcagcactcgaagaccatccttgctgctggcagcatagtctgtt 351
           ||||||||||||| ||||||||||| ||| || |||||| |||||||| | | | |
Db   1204 CCGGTCTCGTGTTCCTTCGCCTGCTTTTGCCCTGCTTGCTGTTTTGTGCTTGTTTGCTGTGCT 1145

Qy   352 tgctgcctggtctgggacacctgtactacagtacttcgttccctttgtagtg 407
           || ||||| |||||| | || | || || ||| || || | | |||||| ||
Db   1144 TGTTGCTTGCTGTCTCGTGTCGTTCTCGGTTCTCGGTTCTCGTTTTGTCGTTTTGTTGTG 1089
```

RESULT 42
AQ665985

```
Qy    306  cgactcgaagacttccttgctgtactccgctggagccagcatagtctgttgctgcctggtctg  365
           | |||||||||| ||| || |||| ||| || ||  || || ||  |||||| ||||| |||||
Db      6  GCAATTGTTGACGTCCTAGNACTNCCCGCTNGCAACCNAGAATCTTTGGACCCTGAGCCC  65

Qy    366  gggacacctgtactacagtacttcgttccctttgtagt  406
           |||||||| ||| || ||| ||  || |  | |||||||
Db     66  CANTCAACAGTGTAGACCATTTCAGCATTTGCTNTAAT  106
```

RESULT 43
AZ233398/c
LOCUS   AZ233398   747 bp   DNA   GSS   14-JUN-2000
DEFINITION   RPCI-23-85A5.TV RPCI-23 Mus musculus genomic clone RPCI-23-85A5,
              DNA sequence.
ACCESSION   AZ233398
VERSION   AZ233398.1 GI:8541444
KEYWORDS   GSS.
SOURCE   house mouse.
```

---

```
LOCUS         AQ665985       539 bp      DNA         GSS        23-JUN-1999
DEFINITION    HS_5351_B2_A10_T7A RPCI-11 Human Male BAC Library Homo sapiens
              genomic clone Plate=927 Col=20 Row=B, DNA sequence.
ACCESSION     AQ665985
VERSION       AQ665985.1  GI:5173753
KEYWORDS      GSS.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 539)
AUTHORS       Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
              Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
              Hood,L.
TITLE         Sequence-tagged connectors: A sequence approach to mapping and
              scanning the human genome
JOURNAL       Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE       99380589
COMMENT       Contact: Mahairas GG, Wallace JC, Hood L
              High Throughput Sequencing Center
              University of Washington
              401 Queen Anne Avenue North, Seattle, WA 98109, USA
              Tel: (206) 616-3618
              Fax: (206) 616-3887
              Email: jwallace@u.washington.edu
              Clones are derived from the human BAC library RPCI-11. For BAC
              library availability, please contact Pieter de Jong
              (pieter@dejong.med.buffalo.edu). Clones may be purchased from
              BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
              or from Resear h Genetics (info@resgen.com). BAC end Web Server:
              http://www.htsc.washington.edu
              Plate: 927 row: B column: 20
              Seq primer: T7
              Class: BAC ends
              High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1..539
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=927 Col=20 Row=B"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT       145 a       145 c       101 g       141 t       7 others
ORIGIN

Query Match        7.3%;    Score 32.6;  DB 231;  Length 539;
Best Local Similarity  55.4%;  Pred. No. 19;
Matches  56;  Conservative  0;  Mismatches  45;  Indels  0;  Gaps  0;

Qy    306  cgactcgaagacttccttgctgtactccgctggagccagcatagtctgttgctgcctggtctg  365
           | |||||||||| ||| || |||| ||| || ||  || || ||  |||||| ||||| |||||
Db      6  GCAATTGTTGACGTCCTAGNACTNCCCGCTNGCAACCNAGAATCTTTGGACCCTGAGCCC  65

Qy    366  gggacacctgtactacagtacttcgttccctttgtagt  406
           |||||||| ||| || ||| ||  || |  | |||||||
Db     66  CANTCAACAGTGTAGACCATTTCAGCATTTGCTNTAAT  106
```

ORGANIZM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 747)
AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
TITLE   Mouse BAC End Sequences from Library RPCI-23
JOURNAL   Unpublished (1999)
COMMENT   Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_intro.html
            Plate: 85 row: A column: 5
            Seq primer: T7
            Class: BAC ends.
FEATURES   Location/Qualifiers
source   1. .747
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="RPCI-23-85A5"
                /clone_lib="RPCI-23"
                /sex="Female"
                /lab_host="DH10B"
                /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                brain genomic DNA was isolated and partially digested
                with a combination of EcoRI and EcoRI Methylase. Size
                selected DNA was cloned into the pBACe3.6 vector at the
                EcoRI sites. The ligation products were transformed into
                DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT   208 a   188 c   202 g   149 t
ORIGIN

Query Match   7.3%;   Score 32.4; DB 240; Length 747;
Best Local Similarity 54.1%; Pred. No. 24;
Matches   66;   Conservative   0;   Mismatches   56;   Indels   0;   Gaps   0;

Qy   67 ctccaagccggacacggctactgtagctggcacggcataatggaaaacagtccctgccct 126
            ||||||| | |||||| || || ||| ||| ||||||||||| | ||| | ||| |
Db   256 CTTCAAGGTTGGCAACTCAGCCCCAAACAGGAACCCACCCTGGAAAACATTGCCTTCGAT 197

Qy   127 tttgagggtcgtctcaaacatgccccctgagacgtcggtgaagccgtagtttgtacaaga 186
            |||||| ||| | || | || |||| | ||||| | | ||| | ||||||
Db   196 TTTGAGGTACTGTTCTGGCATGGCTCTTGGGAGTCCTTAGCCCCGGGTCTTTATGCCAGA 137

Qy   187 ct 188
            ||
Db   136 GT 135

RESULT 44
BF964711/c
LOCUS   BF964711   840 bp   mRNA   EST   22-JAN-2001
DEFINITION 60226781F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4356219 5',
            mRNA sequence.
ACCESSION   BF964711
VERSION   BF964711.1 GI:12331837
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   1 (bases 1 to 840)
TITLE   NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1215 row: k column: 04
            High quality sequence stop: 425.
FEATURES   Location/Qualifiers
source   1. .840
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4356219"
                /clone_lib="NIH_MGC_81"
                /lab_host="DH10B (T1 phage-resistant)"
                /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                (Clontech); Site_1: sfiI (ggccgccctcggcc); Site_2: sfiI
                (ggccattatggcc); 5' and 3' adaptors were used in cloning
                as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
                and 3' adaptor sequence:
                5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                C, or G and N = A, C, G, or T). Average insert size
                1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
                inserts by PCR. This library was enriched for full-length
                clones and was constructed by Clontech Laboratories (Palo
                Alto, CA)."
BASE COUNT   263 a   202 c   234 g   141 t
ORIGIN

Query Match   7.3%;   Score 32.4; DB 171; Length 840;
Best Local Similarity 47.5%; Pred. No. 24;
Matches   96;   Conservative   0;   Mismatches 106;   Indels   0;   Gaps   0;

Qy   200 tggcgtcgatgtgtattccggagccattcgtgctgcgtgttactgtgtagccgcttgatgacttc 259
            |||| ||| ||| || | | |||||| |||| | ||||||| || ||||| | || ||
Db   807 TTGGGTCTCCTGTTGTATAGTCTCCCCTGTGCTGTCCTCTGTGTCGCTCCTTCTCTGTGTTT 748

Qy   260 cttgctgtactccgctggagccatttgagtgctgttaccgtgcagcactcgaagactt 319
            | ||| | | ||| || || || | || ||| | | |||||| | ||| || |
Db   747 TTGCTGCAGCTGCTCTGTGCGTGTTCCTCGCGGGGTGCTTGTGTCGTGTCCTCTGCGGTGGGATAAGT 688

Qy   320 ccttgctgtactcgctggcagcatagtctgtttgtcggtctgtcgggacacctgta 379
            | ||| | | || | | ||| ||| | || |||| || | | | |||| | ||||
Db   687 GCATCCCGTCTCGTTTCGTGTGCTCACGGTGGTTTTCTCTGTGCTGTTTCTCGGTGCTGCTGTGGG 628

Qy   380 ctacagtacttcgtttcccttt 401
            | ||| | || ||| || |||
Db   627 CATTGGTGTGTCTTTTTCTCTCT 606

RESULT 45
AQ175373
LOCUS   AQ175373   437 bp   DNA   GSS   17-OCT-1998
DEFINITION HS_3212_B2_H05_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3212 Col=10 Row=P, DNA sequence.
ACCESSION   AQ175373
VERSION   AQ175373.1 GI:3572740
KEYWORDS   GSS.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 437)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

TITLE   Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
            Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE   99380589
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3212 row: P column: 10
            Class: BAC ends
            High quality sequence stop: 437.
FEATURES   Location/Qualifiers
     source   1..437
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3212 Col=10 Row=P"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-COli DH10B"
BASE COUNT   72 a   144 c   124 g   96 t   1 others
ORIGIN

Query Match   7.3%;   Score 32.2; DB 225; Length 437;
Best Local Similarity   56.0%;   Pred. No. 24;
  Matches   61;   Conservative   0;   Mismatches   48;   Indels   0; Gaps   0;

Qy   302 tgcagcactcgaagacttcctgctgctgtactcgctggcagcatagtctgtttgctgcctgg 361
          | |-||--| | ||||--| || | |||||| ||-|| | |--| | |||
Db   241 TCCGGCGGAGCTGTGTTCGTCGCGCTACTCACTGTCATCGGCCTGCCGTGCCTCAGCCACT 300

Qy   362 tctgggacacctgtgtactacagtacttcgtttccctttgtagtggta 410
          |||||-|||||||| ||| |-||| ||||||-|-|||
Db   301 TCTGGTGACACCTGCTCCGCACATAGTTGCCTTTTCCTTTCAACTGGCA 349

Search completed: September 7, 2001, 17:04:20
Job time: 2444 sec

```
              GenCore version 4.5
       Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       September 7, 2001, 16:16:26 ; Search time 114.83 seconds
                                            (without alignments)
                              2427.835 Million cell updates/sec

Title:        US-09-540-234-1
Perfect score: 444
Sequence:     1 gggtccacggcccttatgt........tgtactcgtctgttcgag 444

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched:     730101 seqs, 31950809 residues

Total number of hits satisfying chosen parameters:    1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    N_Geneseq_0601:*
     1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
     2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
     3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
     4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
     5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
     6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
     7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
     8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
     9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
    10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
    11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
    12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
    13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
    14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
    15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
    16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
    17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
    18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
    19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
    20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
    21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
    22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB Length	ID	Description
1	33.6	7.6	272	21	AAC70962	Single nucleotide
2	33.6	7.6	272	21	AAC70965	Single nucleotide
3	33.6	7.6	272	21	AAC70992	Single nucleotide
4	33.6	7.6	272	21	AAC71013	Single nucleotide
5	33.6	7.6	272	21	AAC71019	Single nucleotide
c 6	33.4	7.5	1894	21	AAA250691	Human prostate der
c 7	33.4	7.5	1894	22	AAC83261	Gland-specific Ets
c 8	33.4	7.5	3317	22	AAC83266	Gland-specific Ets
c 9	32.8	7.4	1713	20	AAA210357	Partial genomic se
10	31.6	7.1	2350	9	AAN82401	A and A' subunits
11	31.6	7.1	2350	12	AAQ14687	Factor XIII subuni
12	31.6	7.1	3425	21	AAC78122	Human cancer assoc
13	31.6	7.1	3905	8	AAN70461	Sequence encoding
14	31.6	7.1	3905	13	AAQ25893	Human Factor XIII
15	31.2	7.0	946	21	AAC41546	Arabidopsis thalia
16	31.2	7.0	1278	21	AAC41498	Arabidopsis thalia
17	31.2	7.0	1440	21	AAC49866	Arabidopsis thalia
18	29.4	6.6	1868	17	AAT39340	DNA sequence for m
c 19	29.4	6.6	21636	21	AAA55966	Human G713 3'-end
c 20	29.2	6.6	1251	20	AAA19436	M. tuberculosis an
c 21	29.2	6.6	1251	20	AAZ19224	M. tuberculosis re
22	29.2	6.6	3515	20	AAX03038	Human IL-1ra BAC c
c 23	29.2	6.6	11901	20	AAX02998	Human IL-1ra BAC c
24	29	6.5	411	21	AAC26317	Human secreted pro
25	29	6.5	488	21	AAC06849	Human secreted pro
26	29	6.5	676	21	AAC77597	Human ORFX ORF3152
27	29	6.5	1193	21	AAC52561	Arabidopsis thalia
28	29	6.5	1889	21	AAA13984	Taxus canadensis g
29	29	6.5	2812	21	AAC59139	Human secreted pro
c 30	28.8	6.5	1889	21	AAC36342	Arabidopsis thalia
c 31	28.8	6.5	10732	21	AAA10594	Gene encoding a su
c 32	28.6	6.4	647	22	AAF74196	DNA encoding eviro
c 33	28.6	6.4	5113	21	AAC83721	Rat laminin 5 cDNA
c 34	28.6	6.4	5264	21	AAF97897	Rat laminin 5 cDNA
35	28.4	6.4	2288	22	AAF97897	Human secreted.pro
c 36	28	6.3	815	22	AAF22709	Human gastric canc
37	28	6.3	50000	21	AAA96363	Polymorphic repeat
c 38	27.8	6.3	1253	20	AAX57500	Rat Ptx3 cDNA. Ra
c 39	27.8	6.3	1392	20	AAX59545	Nucleotide sequenc
40	27.8	6.3	1663	20	AAX13214	Enterococcus faeca
c 41	27.8	6.3	4740	10	AAN91159	DNA sequence downs
c 42	27.8	6.3	6666	12	AAQ10416	Mutant protease ge
c 43	27.8	6.3	6675	12	AAQ10414	Mutant protease ge
c 44	27.8	6.3	6675	12	AAQ10415	Mutant protease ge
c 45	27.8	6.3	6675	12	AAQ10411	Mutant protease ge

ALIGNMENTS

```
RESULT 1
AAC70962
ID   AAC70962 standard; DNA; 272 BP.
XX
AC   AAC70962;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Single nucleotide polymorphism containing sequence #264.
XX
KW   Single nucleotide polymorphism; SNP; human; genetic disease;
KW   disease susceptibility; cardiovascular system; endocrine system;
KW   neurological system; forensic testing; paternity testing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058519-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08440.
XX
PR   31-MAR-1999; 99US-0127248.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Altshuler D, Cargill M, Daley GQ, Ireland JS, Lander ES;
PI   Lipshutz RJ, Patil N, Sklar P;
XX
DR   WPI; 2000-611722/58.
XX
PT   Nucleic acid selected from one of 106 genes comprising single
PT   nucleotide polymorphisms, allele-specific oligonucleotides to the genes
```

```
PT   are useful for phenotypic correlations, forensics, paternity testing,
PT   medicine and genetic analysis -
XX
PS   Claim 1; Fig 5; 214pp; English.
XX
CC   The present invention is concerned with a number of human single
CC   nucleotide polymorphisms (SNPs) which the inventors identified in human
CC   genes. These SNPs can be used in disease diagnosis and prediction of an
CC   individual's susceptibility to disease, in forensic and paternity testing
CC   and in genetic mapping. In particular, the SNPs of the invention can be
CC   used to diagnose susceptibility to diseases of the cardiovascular,
CC   endocrine and neurological systems, such as coronary artery disease,
CC   schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC   diseases.
CC   Note: The degenerate codon within the sequence represents the position
CC   of an SNP, for example the letter S represents a polymorphism where the
CC   nucleotide may be C or G.
XX
SQ   Sequence 272 BP; 65 A; 56 C; 74 G; 76 T; 1 other;

Query Match              7.6%;   Score 33.6; DB 21;  Length 272;
Best Local Similarity   59.4%;   Pred. No. 0.15;
Matches   57; Conservative    0; Mismatches   39; Indels    0; Gaps    0;

Qy  135 tcgtctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
        || |||| ||| | ||| ||||| | |||| ||||||||||||||||| || ||||||
Db  108 tcctttaaaagaaaccctgcgaaatgtctgggtacacctggatggtcctggagtaacaag 167

Qy  195 gccaatggcgtcgatgttatcggagccattcgtgtg 230
        ||||||| | |||||| | |||||| ||| || ||
Db  168 accaatgaagaagatgttccggtaaaccttgggttg 203

RESULT  2
AAC70965
ID   AAC70965 standard; DNA; 272 BP.
XX
AC   AAC70965;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Single nucleotide polymorphism containing sequence #265.
XX
KW   Single nucleotide polymorphism; SNP; human; genetic disease;
KW   disease susceptibility; cardiovascular system; endocrine system;
KW   neurological system; forensic testing; paternity testing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058519-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08440.
XX
PR   31-MAR-1999;  99US-0127248.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Altshuler D,  Cargill M,  Daley GQ,  Ireland JS,  Lander ES;
PI   Lipshutz RJ,  Patil N,  Sklar P;
XX
DR   WPI; 2000-611722/58.
XX
PT   Nucleic acid selected from one of 106 genes comprising single
PT   nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT   are useful for phenotypic correlations, forensics, paternity testing,
PT   medicine and genetic analysis -
XX
PS   Claim 1; Fig 5; 214pp; English.
XX
```

```
CC   The present invention is concerned with a number of human single
CC   nucleotide polymorphisms (SNPs) which the inventors identified in human
CC   genes. These SNPs can be used in disease diagnosis and prediction of an
CC   individual's susceptibility to disease, in forensic and paternity testing
CC   and in genetic mapping. In particular, the SNPs of the invention can be
CC   used to diagnose susceptibility to diseases of the cardiovascular,
CC   endocrine and neurological systems, such as coronary artery disease,
CC   schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC   diseases.
CC   Note: The degenerate codon within the sequence represents the position
CC   of an SNP, for example the letter S represents a polymorphism where the
CC   nucleotide may be C or G.
XX
SQ   Sequence 272 BP; 66 A; 55 C; 74 G; 76 T; 1 other;

Query Match              7.6%;   Score 33.6; DB 21;  Length 272;
Best Local Similarity   59.4%;   Pred. No. 0.15;
Matches   57; Conservative    0; Mismatches   39; Indels    0; Gaps    0;

Qy  135 tcgtctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
        || |||| ||| | ||| ||||| | |||| ||||||||||||||||| || ||||||
Db  108 tcctttaaaagaaaccctgcgaaatgtctgggtacacctggatggtcctggagtaacaag 167

Qy  195 gccaatggcgtcgatgttatcggagccattcgtgtg 230
        ||||||| | |||||| | |||||| ||| || ||
Db  168 accaatgaagaagatgttccggtaaaccttgggttg 203

RESULT  3
AAC70992
ID   AAC70992 standard; DNA; 272 BP.
XX
AC   AAC70992;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Single nucleotide polymorphism containing sequence #274.
XX
KW   Single nucleotide polymorphism; SNP; human; genetic disease;
KW   disease susceptibility; cardiovascular system; endocrine system;
KW   neurological system; forensic testing; paternity testing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058519-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08440.
XX
PR   31-MAR-1999;  99US-0127248.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Altshuler D,  Cargill M,  Daley GQ,  Ireland JS,  Lander ES;
PI   Lipshutz RJ,  Patil N,  Sklar P;
XX
DR   WPI; 2000-611722/58.
XX
PT   Nucleic acid selected from one of 106 genes comprising single
PT   nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT   are useful for phenotypic correlations, forensics, paternity testing,
PT   medicine and genetic analysis -
XX
PS   Claim 1; Fig 5; 214pp; English.
XX
CC   The present invention is concerned with a number of human single
CC   nucleotide polymorphisms (SNPs) which the inventors identified in human
CC   genes. These SNPs can be used in disease diagnosis and prediction of an
CC   individual's susceptibility to disease, in forensic and paternity testing
CC   and in genetic mapping. In particular, the SNPs of the invention can be
```

```
CC     used to diagnose susceptibility to diseases of the cardiovascular,
CC     endocrine and neurological systems, such as coronary artery disease,
CC     schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC     diseases.
CC     Note: The degenerate codon within the sequence represents the position
CC     of an SNP, for example the letter S represents a polymorphism where the
CC     nucleotide may be C or G.
XX
SQ     Sequence 272 BP; 66 A; 56 C; 74 G; 75 T; 1 other;

Query Match             7.6%;  Score 33.6;  DB 21;  Length 272;
Best Local Similarity  59.4%;  Pred. No. 0.15;
Matches   57;  Conservative    0;  Mismatches  39;  Indels    0;  Gaps    0;

Qy      135 tcgtctcaaacatgccctcgagacgtcggtgaagcctagtttgtacaagactaacaac 194
            || |-|| ||| ||| ||-||||||-| |||||||-|- ||||||-|||-||| || ||||
Db      108 tcctttaaaagaaacctgcgaaatgtctggtcacctggtacacctggagtaacaag 167

Qy      195 gccaatggcgtcgatgttatcggagccattcgtgtg 230
            |||||| |-||-|||||||||| ||-| -|-|| ||
Db      168 accaatgaagaagatgttccggtaaccctgggttg 203

RESULT 4
AAC71013
ID   AAC71013 standard; DNA; 272 BP.
XX
AC   AAC71013;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Single nucleotide polymorphism containing sequence #281.
XX
KW   Single nucleotide polymorphism; SNP; human; genetic disease;
KW   disease susceptibility; cardiovascular system; endocrine system;
KW   neurological system; forensic testing; paternity testing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058519-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08440.
XX
PR   31-MAR-1999;  99US-0127248.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Altshuler D,  Cargill M,  Daley GQ,  Ireland JS,  Lander ES;
PI   Lipshutz RJ,  Patil N,  Sklar P;
XX
DR   WPI; 2000-611722/58.
XX
PT   Nucleic acid selected from one of 106 genes comprising single
PT   nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT   are useful for phenotypic correlations, forensics, paternity testing,
PT   medicine and genetic analysis -
XX
PS   Claim 1; Fig 5; 214pp; English.
XX
CC   The present invention is concerned with a number of human single
CC   nucleotide polymorphisms (SNPs) which the inventors identified in human
CC   genes. These SNPs can be used in disease diagnosis and prediction of an
CC   individual's susceptibility to disease, in forensic and paternity testing
CC   and in genetic mapping. In particular, the SNPs of the invention can be
CC   used to diagnose susceptibility to diseases of the cardiovascular,
CC   endocrine and neurological systems, such as coronary artery disease,
CC   schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC   diseases.
CC   Note: The degenerate codon within the sequence represents the position
```

```
CC     of an SNP, for example the letter S represents a polymorphism where the
CC     nucleotide may be C or G.
XX
SQ     Sequence 272 BP; 66 A; 56 C; 73 G; 76 T; 1 other;

Query Match             7.6%;  Score 33.6;  DB 21;  Length 272;
Best Local Similarity  59.4%;  Pred. No. 0.15;
Matches   57;  Conservative    0;  Mismatches  39;  Indels    0;  Gaps    0;

Qy      135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
            || |-|| ||| ||| ||-||||||-| |||||||-|- ||||||-|||-||| || ||||
Db      108 tcctttaaaagaaacctgcgaaatgtctggtcacctggtacacctggagtaacaag 167

Qy      195 gccaatggcgtcgatgttatcggagccattcgtgtg 230
            |||||| |-||-|||||||||| ||-| -|-|| ||
Db      168 accaatgaagaagatgttccggtaaccctgggttg 203

RESULT 5
AAC71019
ID   AAC71019 standard; DNA; 272 BP.
XX
AC   AAC71019;
XX
DT   09-FEB-2001  (first entry)
XX
DE   Single nucleotide polymorphism containing sequence #283.
XX
KW   Single nucleotide polymorphism; SNP; human; genetic disease;
KW   disease susceptibility; cardiovascular system; endocrine system;
KW   neurological system; forensic testing; paternity testing; ds.
XX
OS   Homo sapiens.
XX
PN   WO200058519-A2.
XX
PD   05-OCT-2000.
XX
PF   30-MAR-2000; 2000WO-US08440.
XX
PR   31-MAR-1999;  99US-0127248.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
PA   (AFFY-) AFFYMETRIX INC.
XX
PI   Altshuler D,  Cargill M,  Daley GQ,  Ireland JS,  Lander ES;
PI   Lipshutz RJ,  Patil N,  Sklar P;
XX
DR   WPI; 2000-611722/58.
XX
PT   Nucleic acid selected from one of 106 genes comprising single
PT   nucleotide polymorphisms, allele-specific oligonucleotides to the genes
PT   are useful for phenotypic correlations, forensics, paternity testing,
PT   medicine and genetic analysis -
XX
PS   Claim 1; Fig 5; 214pp; English.
XX
CC   The present invention is concerned with a number of human single
CC   nucleotide polymorphisms (SNPs) which the inventors identified in human
CC   genes. These SNPs can be used in disease diagnosis and prediction of an
CC   individual's susceptibility to disease, in forensic and paternity testing
CC   and in genetic mapping. In particular, the SNPs of the invention can be
CC   used to diagnose susceptibility to diseases of the cardiovascular,
CC   endocrine and neurological systems, such as coronary artery disease,
CC   schizophrenia, cancer, autoimmune diseases, Alzheimer's and Parkinson's
CC   diseases.
CC   Note: The degenerate codon within the sequence represents the position
CC   of an SNP, for example the letter S represents a polymorphism where the
CC   nucleotide may be C or G.
XX
SQ     Sequence 272 BP; 66 A; 55 C; 74 G; 76 T; 1 other;
```

Query Match   7.6%; Score 33.6; DB 21; Length 272;
Best Local Similarity   59.4%; Pred. No. 0.15;
Matches   57; Conservative   0; Mismatches   39; Indels   0; Gaps   0;

Qy 135 tcgtctcaaacatgccctcgagacgtcgtgagcgtgagcgtgagcgtggtgtacaagactaacaac 194
            || | ||| | || ||| | || | | | | || |||||
Db 108 tcctttaaagaaacctgcgaaatgtctggtcacctgatggtcctggagtaacaag 167

Qy 195 gccaatggcgtcgatgttatcggaggccattcgtgtg 230
           ||||| | ||||| | ||||| || | ||
Db 168 accaatgaagagatgttccggtaaaccctgggttg 203

RESULT 6
AA250691/c
ID   AA250691 standard; cDNA; 1894 BP.
XX
AC   AA250691;
XX
DT   23-MAY-2000 (first entry)
XX
DE   Human prostate derived Ets factor cDNA.
XX
KW   Prostate Derived Ets factor; PDEF; human; chromosome 6p21.3; cancer;
KW   loss of heterozygosity; chromosomal translocation; linkage analysis;
KW   cytostatic; cardiant; immunosuppressive; cerebroprotective; fungicide;
KW   antibacterial; vulnerary; neuroprotective; antiparkinsonian; nootropic;
KW   anabolic; antiinflammatory; anorectic; hybridisation probe; forensic;
KW   tumour marker; diagnosis; treatment; prostate cancer; blood coagulation;
KW   autoimmune disorder; haematopoietic; immune/nervous system; stroke;
KW   neoplasm; microbial infection; tissue regeneration; heart attack;
KW   scarring; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   416..1423
FT   /*tag= a
FT   /product= "Human prostate derived Ets factor"
FT   /note= "Expressed in prostate epithelium"
XX
PN   WO200006589-A1.
XX
PD   10-FEB-2000.
XX
PF   02-AUG-1999; 99WO-US17470.
XX
PR   31-JUL-1998; 98US-0126945.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Libermann TA, Oettgen JP, Kunsch CA, Endress GA, Rosen CA;
XX
DR   WPI; 2000-195255/17.
DR   P-PSDB: AAY45003.
XX
PT   Novel prostate derived polypeptide, polynucleotide useful for
PT   diagnosis, prevention and treatment of prostate cancer, autoimmune
PT   disorders, microbial infections and also as food additive or
PT   preservative -
XX
PS   Claim 4; Fig 1; 132pp; English.
XX
CC   The present cDNA sequence encodes a human prostate derived Ets factor
CC   (PDEF), isolated from human prostate carcinoma cDNA library
CC   (ATCC No. 203072). PDEF gene expression is highest in tissues involved in
CC   androgen sensitivity, with enriched expression in prostate epithelium
CC   cells. PDEF gene is mapped to the human chromosome 6p21.3
CC   region that is associated with loss of heterozygosity and chromosomal
CC   translocations in various human cancers. PDEF has cytostatic, cardiant,
CC   immunosuppressive, cerebroprotective, fungicide, antibacterial,

CC   vulnerary, neuroprotective, antiparkinsonian, nootropic, anabolic,
CC   antiinflammatory and anorectic activity. PDEF polynucleotides are useful
CC   in linkage analysis as markers, as hybridisation probes for differential
CC   identification of the tissues or cell types and as polymorphic markers
CC   for forensic purposes. PDEF is useful as prostate-specific tumour marker
CC   for the diagnosis and treatment of prostate cancer. PDEF sequences are
CC   useful for treating autoimmune disorders, haematopoietic, blood
CC   coagulation, immune and nervous system disorders, hyperproliferative
CC   disorders like, neoplasms and microbial infections, heart attacks,
CC   stroke, scarring and for tissue regeneration. They are also useful as
CC   food additives or preservatives.
XX
SQ   Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

Query Match   7.5%; Score 33.4; DB 21; Length 1894;
Best Local Similarity   58.6%; Pred. No. 0.43;
Matches   58; Conservative   0; Mismatches   41; Indels   0; Gaps   0;

Qy 290 tgctgtttaccgtgcagcactcgaagacttcctgcgtactcgctggcagcatagtctg 349
            |||| ||| | || | || | | || | ||
Db 443 TGCTCAGACCCGGCGTGCGCGTGCCCATGCCGTGCGTGCTGCGTGTTGGCGTCGCGGCGTCGTGTCTG 384

Qy 350 tttgctgcctgtctgggacacctgtgtactacagtac 388
            ||||||||| ||| ||||||| | |||| |
Db 383 TTAGCTGCCTGGTGCCCAGGGAGCTGTCTGCTGCAGTGC 345

RESULT 7
AAC83261/c
ID   AAC83261 standard; DNA; 1894 BP.
XX
AC   AAC83261;
XX
DT   16-MAR-2001 (first entry)
XX
DE   Gland-specific Ets transcription factor (GSEF) nucleotide sequence.
XX
KW   Transcription factor; gland-specific Ets transcription factor; GSEF;
KW   metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW   leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ds.
XX
OS   Homo sapiens.
XX
PN   WO200070092-A1.
XX
PD   23-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US13173.
XX
PR   14-MAY-1999; 99US-0134112.
XX
PA   (CHIR) CHIRON CORP.
XX
PI   Kaufmann J, Xin H, Harrowe G;
XX
DR   WPI; 2001-041019/05.
DR   P-PSDB; AAB49628.
XX
PT   Detecting metastatic and potential metastatic cancerous cells, useful
PT   for diagnosing, prognosing, grading and staging of cancers by detecting
PT   gland-specific Ets transcription factor (GSEF) gene product in a biological
PT   sample from a cell -
XX
PS   Disclosure; Page 81-83; 95pp; English.
XX
CC   This invention relates to a method for the detection and determination of
CC   the metastatic potential of a cell. The method comprises detecting a
CC   gland-specific Ets transcription factor (GSEF) gene product in a test
CC   sample. Detection of a GSEF gene product in the test sample in amount
CC   lower than that in a normal cell, is indicative of a cell with high
CC   metastatic potential. The method is useful for determining the metastatic
CC   potential of a cell, for the diagnosis and prognosis of cancer as well as

Left column:

CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer-associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the DNA encoding GSEF.
XX
SQ Sequence 1894 BP; 368 A; 653 C; 571 G; 302 T; 0 other;

Query Match   7.5%; Score 33.4; DB 22; Length 1894;
Best Local Similarity 58.6%; Pred. No. 0.43;
Matches 58; Conservative   0; Mismatches 41; Indels   0; Gaps   0;

Qy 290 tgctgtttaccgtgcagcactcgaagacttcctgctgtactcgctggcagcatagtctg 349
Db 443 TGCTCAGACCCGGCTGGCGCTGCCCATGCCGCTGCTGTTTGGGCTGGCGGCTGTGTCTG 384

Qy 350 tttgctgcctggtctgggacacctgtgtactacagtac 388
Db 383 TTAGCTGCCTGGTGCCCAGGAGCTGTCTGCTGCAGTGC 345

RESULT 8
AAC83266/c
ID AAC83266 standard; DNA; 3317 BP.
XX
AC AAC83266;
XX
DT 16-MAR-2001 (first entry)
XX
DE Gland-specific Ets transcription factor (GSEF) cDNA sequence.
XX
KW Transcription factor; gland-specific Ets transcription factor; GSEF;
KW metastatic potential; cancer; tumour; metastasis; breast; prostate;
KW leukaemia; lymphoma; sarcoma; melanoma; chromosome 6p21.1-6p21.3; ss.
XX
OS Homo sapiens.
XX
PN WO200070092-A1.
XX
PD 23-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US13173.
XX
PR 14-MAY-1999; 99US-0134112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Kaufmann J, Xin H, Harrowe G;
XX
DR WPI; 2001-041019/05.
DR P-PSDB; AAB49633.
XX
PT Detecting metastatic and potential metastatic cancerous cells, useful
PT for diagnosing, prognosing, grading and staging of cancers by detecting
PT gland-specific Ets transcription factor gene product in a biological
PT sample from a cell -
XX
PS Disclosure; Fig 2; 95pp; English.
XX
CC This invention relates to a method for the detection and determination of
CC the metastatic potential of a cell. The method comprises detecting a
CC gland-specific Ets transcription factor (GSEF) gene product in a test

Right column:

CC sample. Detection of a GSEF gene product in the test sample in amount
CC lower than that in a normal cell, is indicative of a cell with high
CC metastatic potential. The method is useful for determining the metastatic
CC potential of a cell, for the diagnosis and prognosis of cancer as well as
CC grading and staging of cancers by detecting GSEF expression in a
CC biological test sample. The method may also be used to monitor patients
CC having a predisposition to develop a particular cancer. GSEF polypeptides
CC are useful for producing antibodies, in cancer diagnosis, prognosis,
CC grading, staging and management of breast and prostate tumours, and in
CC detecting polymorphisms in the sequence. GSEF genes and proteins are also
CC useful in gene therapy. GSEF gene product expression levels can be used
CC in conjunction with any tissue in which an alteration in GSEF gene
CC product expression levels is associated with development of a
CC cancer-associated phenotype. Cancers, which can be monitored include
CC cancers of the prostate, cervix, lung and colon, melanomas, colorectal
CC adenocarcinomas, Wilm's tumour, retinoblastoma, sarcomas, myosarcomas,
CC lung carcinomas, leukaemia, and lymphomas. The GSEF gene is located on
CC human chromosome 6, specifically at 6p21.1-6p21.3. The present sequence
CC represents the cDNA encoding GSEF.
XX
SQ Sequence 3317 BP; 710 A; 1026 C; 970 G; 611 T; 0 other;

Query Match   7.5%; Score 33.4; DB 22; Length 3317;
Best Local Similarity 58.6%; Pred. No. 0.56;
Matches 58; Conservative   0; Mismatches 41; Indels   0; Gaps   0;

Qy 290 tgctgtttaccgtgcagcactcgaagacttcctgctgtactcgctggcagcatagtctg 349
Db 1842 TGCTCAGACCCGGCTGGCGCTGCCCATGCCGCTGCTGTTTGGGCTGGCGGCTGTGTCTG 1783

Qy 350 tttgctgcctggtctgggacacctgtgtactacagtac 388
Db 1782 TTAGCTGCCTGGTGCCCAGGAGCTGTCTGCTGCAGTGC 1744

RESULT 9
AAZ10357/c
ID AAZ10357 standard; DNA; 1713 BP.
XX
AC AAZ10357;
XX
DT 15-NOV-1999 (first entry)
XX
DE Partial genomic sequence of the murine DLC-1 gene.
XX
KW DLC-1 gene; tumor suppressor gene; liver cancer; gene therapy.
KW hepatocellular carcinoma; prostate cancer; colon cancer; rectum cancer;
KW breast cancer; adenocarcinoma; carcinogenesis; ss.
XX
OS Mus sp.
XX
PN WO943812-A2.
XX
PD 02-SEP-1999.
XX
PF 25-FEB-1999; 99WO-US04164.
XX
PR 25-FEB-1998; 98US-0075952.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Popescu NC, Thorgeirsson SS, Yuan B;
XX
DR WPI; 1999-540590/45.
XX
PT New nucleic acid representing the human deleted in liver cancer-1
PT gene, used for diagnosis and gene therapy of cancer
XX
PS Example 7; Page 67-68; 85pp; English.
XX
CC AAZ10357-62 represent partial genomic sequences of the murine DLC-1
CC gene. This gene is a putative tumor suppressor gene which is frequently

```
CC   deleted in liver cancer cells. Detecting deletion of this gene, or
CC   absence of the expressed protein, indicates increased susceptibility
CC   to cancer, or presence of cancer (particularly hepatocellular carcinoma;
CC   cancer of prostate, colon/rectum or breast, or adenocarcinoma). The
CC   DLC-1 gene, or its cDNA, can be used in gene therapy to replace lost
CC   gene function, specifically for treating cancer; and to generate
CC   knockout transgenic animals (as in vivo models of carcinogenesis).
XX
SQ   Sequence 1713 BP: 433 A; 485 C; 451 G; 326 T; 18 other;

Query Match      7.4%;   Score 32.8; DB 20;   Length 1713;
Best Local Similarity   53.0%;   Pred. No. 0.65;
Matches   70;  Conservative   0;  Mismatches   62;  Indels   0;  Gaps   0;

Qy  223 ttcgtgtgctgtttactgtgtagcgcttgatgacttcctttgctactccgcttggagcc 282
        | ||| || ||||| ||| ||| || ||||| | ||||| |||| ||| || |||||
Db  697 TCCGGTCCTGGTGACCGGCGGGCTGGGTGTGCCGCCGCCGCTGGTGTCCTGATTGGCTGC 638

Qy  283 atttgagtgctgtttaccgtgcagcactcgaagaacttccttgctgtactcgcttggcagca 342
        |  ||||| || ||||||||| | |||| |||| ||||||||| | |||| ||||| ||||
Db  637 TGCTTGCTGCTGGTCTGGGTGAAGTTAGACACGGCTCCTTTCCGGTACCATGGGCACGT 578

Qy  343 tagtctgtttgc 354
        |  ||| |||||
Db  577 TGAATGTGGTTGC 566
```

RESULT 10
AAN82401
```
ID   AAN82401 standard; DNA; 2350 BP.
XX
AC   AAN82401;
XX
DT   26-NOV-1990   (first entry)
XX
DE   A and A' subunits of human Factor XIII.
XX
KW   A sub-unit; factor XIII; blood clot stabilisation; fibrin polymer;
KW   crosslink; scleroderma; haemorrhage; ulcerative colitis; ss.
XX
OS   Homo sapiens.
XX
FH   Key               Location/Qualifiers
FT   CDS               91..2284
FT                     /*tag= a
FT                     /product=mature a subunit
FT                     202..2284
FT                     /*tag= b
FT                     /product=mature a' subunit
XX
PN   AU878694-A.
XX
PD   31-MAR-1988.
XX
PF   21-SEP-1987;  87AU-0078694.
XX
PR   19-SEP-1986;  86US-0909512.
XX
PA   (ZYMO-) ZYMOGENETICA INC.
XX
PI   Davie E,  Seale RL,  Ichinose A,  Holly JA,  Parker GE;
XX
DR   WPI; 1988-140637/21.
DR   P-PSDB;  AAP82920.
XX
PT   New DNA sequences encoding sub-units of factor 13 - and corresponding
PT   expression vectors and transformed host cells.
XX
PS   Disclosure; ; p; English.
XX
CC   The carboxyl-terminal Met (nucleotides 2281-2283) is followed by
CC   a stop codon (TGA), 1535 bp noncoding sequence, and a potential
```

```
CC   polyadenylation or processing signal of AATAAA. The polyadenylation
CC   sequence was located 14 nucleotides upstream from the poly(A) tail
CC   of 10 nucleotides. The poly(A) tail was present only in a second
CC   cDNA clone, designated lambdaHFXIIIa3.82.
CC   A difference in the nucleotide sequence for the a subunit was found
CC   at three positions when comparison of the cDNA inserts was made in
CC   regions where overlapping sequences were obtained. Nucleotides
CC   2038, 2041, and 2727 contained A, C, and T resp. in
CC   lambdaHFXIIIa3.77, while lambdaHFXIIIa3.82 contained G, C, and A
CC   in the same positions. These differences results in a change in two
CC   amino acids (Ile 680 and Gln 681 to Val and Glu), and could
CC   represent a polymorphism that contributes to the microheterogeneity
CC   in the a subunit of factor XIII (P.G. Board and M. Coggan,
CC   Hum.Genet. 59:135-136).
CC   When cultured the host cells will produce the polypeptide which
CC   can be assembled to factor 13, which stabilises blood clots by
CC   crosslinking fibrin polymers. Factor 13 is useful therapeutically, e.g.
CC   in cases of scleroderma, haemorrhage, ulceratice colitis etc., and
CC   can be prepd. in large amts. without risk of viral contamination.
CC   See also AAN82402 and AAN82403.
XX
SQ   Sequence 2350 BP: 645 A; 553 C; 620 G; 532 T; 0 other;

Query Match      7.1%;   Score 31.6; DB 9;   Length 2350;
Best Local Similarity   62.8%;   Pred. No. 1.9;
Matches   49;  Conservative   0;  Mismatches   29;  Indels   0;  Gaps   0;

Qy  135 tcgtccaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaaggactaacaac 194
        ||||||| ||| ||||| ||||||| |||||||| |||| ||||||||  ||||  ||||||
Db  2052 tcctttaaaagaaccctgcgaaatgtctcggtgtacacctgatggtcctggagtaacag 2111

Qy  195 gccaatggcgtcgatgtt 212
        ||||| ||||||||||
Db  2112 accaatgaagaagatgtt 2129
```

RESULT 11
AAQ14687
```
ID   AAQ14687 standard; cDNA; 2350 BP.
XX
AC   AAQ14687;
XX
DT   14-FEB-1992   (first entry)
XX
DE   Factor XIII subunit a.
XX
KW   Factor XIII; subunit; antibody; diagnosis; thrombosis; mutant; ss.
XX
OS   Homo sapiens.
XX
FH   Key               Location/Qualifiers
FT   CDS               91..886
FT                     /*tag=  a
FT                     /label= factor_XIII
FT                     /note= "mature a subunit"
FT   mat_peptide       202..811
FT                     /*tag=  b
FT                     /note= "mature a'subunit"
XX
PN   WO9116931-A.
XX
PD   14-NOV-1991.
XX
PF   09-MAY-1991;  91WO-US03212.
XX
PR   18-MAY-1990;  90US-0525556.
PR   10-MAY-1990;  90US-0521805.
XX
PA   (ZYMO-) ZYMOGENETICS INC.
XX
PI   Bishop PD;
XX
```

```
DR    WPI; 1991-353537/48.
DR    P-PSDB; AAR14376.
XX
PT    New diagnostic compsns. contg. factor XIII or derivs. -
PT    are coupled to radioisotope or paramagnetic cpd. and are useful
PT    for detecting thrombosis in high risk patients
XX
PS    Disclosure; Fig 1; 51pp; English.
XX
CC    A cDNA library from human placental mRNA was screened with anti-
CC    factor XIII antibody and then with a specific oligonucleotide.
CC    The sequence of one clone, represented here, encodes the entire a
CC    subunit. The C-terminal Met is followed by a stop codon (TGA), 1535
CC    bp of noncoding sequence, and a potential polyadenylation sequence
CC    or processing signal of AATAAA (the sequence of the noncoding
CC    sequence and poly(A) signal are not reproduced). The polyadenylation
CC    sequence was located 14 nucleotides upstream from the poly(A) tail of
CC    10 nucleotides. The poly(A) tail was present only in a second clone,
CC    designated lambdaHFXIIIa3.82.
CC    Mutant forms of factor XIII were made by in vitro site-directed
CC    mutagenesis, esp. ss pRS202 templates were modified by the single
CC    primer method, e.g. for TGC-Cys-314 to TCT-Ser exchange.
CC    See also AAQ14687-88.
XX
SQ    Sequence 2350 BP; 645 A; 556 C; 618 G; 531 T; 0 other;

Query Match             7.1%;    Score 31.6;  DB 12;    Length 2350;
Best Local Similarity  62.8%;    Pred. No. 1.9;
Matches    49;  Conservative    0;  Mismatches   29;  Indels   0;  Gaps   0;

Qy    135 tcgtctcaaacatgcccctcgagacgtcggtgaagccgtagttgtacaagactaacaac 194
          ||| || |||| ||| ||| ||| |||| |||||||| |||| || ||| | ||||||
Db   2052 tcctttaaaagaaaccctgcgaaatgtctgggtacacctgatggtcctggagtaacaag 2111

Qy    195 gccaatggcgtcgatgtt 212
          ||||| | ||||||
Db   2112 accaatgaagaagatgtt 2129

RESULT 12
AAC78122
ID    AAC78122 standard; cDNA; 3425 BP.
XX
AC    AAC78122;
XX
DT    08-FEB-2001  (first entry)
XX
DE    Human cancer associated gene sequence SEQ ID NO:516.
XX
KW    Human; cancer associated gene; cancer antigen; detection; cancer;
KW    diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW    antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW    antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW    dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW    vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW    immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW    allergic reaction; graft versus host disease; organ rejection;
KW    haemostatic; thrombolytic; cardiovascular disorder; infection;
KW    neurological disease; drug screening; ss.
XX
OS    Homo sapiens.
XX
PN    WO200055350-A1.
XX
PD    21-SEP-2000.
XX
PF    08-MAR-2000; 2000WO-US05882.
XX
PR    12-MAR-1999;  99US-0124270.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
```

```
PI    Rosen CA,  Ruben SM;
XX
DR    WPI; 2000-587533/55.
DR    P-PSDB; AAB43913.
XX
PT    Novel isolated nucleic acids comprising sequences encoding peptides
PT    useful for treating or diagnosing e.g. cancer -
XX
PS    Claim 1; Page 1046-1047; 2352pp; English.
XX
CC    AAC77607 to AAC78448 encode the human cancer associated proteins given
CC    in AAB43398 to AAB44239. The proteins can have activities based on the
CC    tissues and cells the genes are expressed in. Example of activities
CC    include: cytostatic; proliferative; vulnerary; immunomodulator;
CC    antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC    antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC    dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC    nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC    polynucleotides and polypeptides can be used for preventing, treating or
CC    ameliorating medical conditions and diagnosing pathological conditions.
CC    Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC    the present invention may be used to treat immune disorders by activating
CC    or inhibiting the proliferation, differentiation or mobilisation of
CC    immune cells, to treat disorders of haematopoietic cells, autoimmune
CC    disorders, allergic reactions, graft versus host disease and organ
CC    rejection, modulate haemostatic or thrombolytic activity, modulate
CC    inflammation, cancers, cardiovascular disorders, neurological disease and
CC    bacterial or viral infections. The peptides, nucleotides, antibodies,
CC    agonists and antagonists may be also be used in drug screens. AAC78449 to
CC    AAC78457 and AAB44240 represent sequences used in the exemplification of
CC    the present invention.
XX
SQ    Sequence 3425 BP; 1001 A; 760 C; 796 G; 865 T; 3 other;

Query Match             7.1%;    Score 31.6;  DB 21;    Length 3425;
Best Local Similarity  62.8%;    Pred. No. 2.3;
Matches    49;  Conservative    0;  Mismatches   29;  Indels   0;  Gaps   0;

Qy    135 tcgtctcaaacatgcccctcgagacgtcggtgaagccgtagttgtacaagactaacaac 194
          ||| || |||| ||| ||| ||| |||| |||||||| |||| || ||| | ||||||
Db   1607 tcctttaaaagaaaccctgcgaaatgtctgggtacacctgatggtcctggagtaacaag 1666

Qy    195 gccaatggcgtcgatgtt 212
          ||||| | ||||||
Db   1667 accaatgaagaagatgtt 1684

RESULT 13
AAN70461
ID    AAN70461 standard; cDNA; 3905 BP.
XX
AC    AAN70461;
XX
DT    21-MAY-1991  (first entry)
XX
DE    Sequence encoding human factor XIIIa.
XX
KW    Diagnosis; antigen; anti-factor XIIIa antibody; ss.
XX
OS    Homo sapiens.
XX
FH    Key              Location/Qualifiers
FT    CDS              85..2283
FT                     /tag= a
FT    misc_feature     764..830
FT                     /tag= b
FT                     /note= "66mer probe binds here"
FT    misc_feature     1506..1526
FT                     /tag= c
FT                     /note= "20mer probe binds here"
XX
PN    EP236978-A.
XX
```

Left column:

```
XX
PD   16-SEP-1987.
XX
PF   06-MAR-1987;   87EP-0103222.
XX
PR   26-JUN-1986;   86DE-3621371.
PR   12-MAR-1986;   86DE-3608280.
XX
PA   (BEHW ) BEHRINGWERKE AG.
XX
PI   Grundmann U, Amann E, Zettlmeissl G;
XX
DR   WPI; 1987-258275/37.
DR   P-PSDB; AAP70293.
XX
PT   New DNA sequence coding for factor 13A and expressed proteins -
PT   useful as diagnostic reagents and for producing antibodies
XX
PS   Claim 1; Table 3, pp16-20; 30pp; German.
XX
CC   Human placental cDNA gene bank was screened by hybridisation with two
CC   synthetic oligonucleotides, corresponding to partial AA sequences of
CC   factor XIIIa (AAN70460, AAN70465). AAN70461 gives the coding strand
CC   sequence of clones lambda-gt10-11 and lambda-gt10-12.
XX
SQ   Sequence 3905 BP; 1175 A; 865 C; 898 G; 967 T; 0 other;

Query Match             7.1%;  Score 31.6;  DB 8;  Length 3905;
  Best Local Similarity  62.8%;  Pred. No. 2.5;
  Matches    49;  Conservative    0;  Mismatches    29;  Indels    0;  Gaps    0;

Qy      135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
              || ||| ||| ||  ||| ||| |||  ||||| | ||| |||| || ||  ||||||
Db     2049 tcctttaaagaaacctcgcgaaatgtctgggtacacctggatggtcctggagtaacaag 2108

Qy      195 gccaatggcgtcgatgtt 212
              ||||| | | |||||
Db     2109 accaatgaagaagatgtt 2126

RESULT 14
AAQ25893
ID   AAQ25893 standard; cDNA; 3905 BP.
XX
AC   AAQ25893;
XX
DT   08-JAN-1993 (first entry)
XX
DE   Human Factor XIII coding sequence.
XX
KW   Factor 13; Factor XIIIa; plasma-transglutaminase; fibrinoligase;
KW   fibrin stabilising factor; blood clotting; coagulation; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             85..2283
FT                   /*tag= a
FT                   /product= Factor_XIII
XX
PN   EP494702-A.
XX
PD   15-JUL-1992.
XX
PF   06-MAR-1987;   87EP-0105735.
XX
PR   12-MAR-1986;   86DE-3608280.
PR   26-JUN-1986;   86DE-3621371.
XX
PA   (BEHW ) BEHRINGWERKE AG.
XX
PI   Amann E, Grundmann U, Zettlmeissl G;
```

Right column:

```
XX
DR   WPI; 1992-235765/29.
DR   P-PSDB; AAR25385.
XX
PT   Pure recombinant protein with factor XIII activity - useful in
PT   diagnosis and treatment of factor XIII defects
XX
PS   Example 1; Page 16-20; 30pp; German.
XX
CC   Mature human placenta RNA was isolated and converted to ds cDNA.
CC   A cDNA library was then prepared in Lambda phage gt10. The library
CC   was screened with two probes (see AAQ25891 and AAQ32258) based on the
CC   sequences of peptide fragments from Factor XIII. One clone
CC   (lambda gt10-12) hybridised with both probes and was found to
CC   contain a 1704 bp Factor XIIIa sequence with an internal EcoRI site.
CC   The two EcoRI fragments generated from the insert sequence (540bp and
CC   1164bp, respectively) were labelled and used to screen the cDNA
CC   library for clones extending at both ends compared to lambda gt10-12.
CC   The full-length sequence coding for the inactive Factor XIII
CC   polypeptide was constructed from the sequences of overlapping clones.
CC   See also AAQ25328-9.
XX
SQ   Sequence 3905 BP; 1175 A; 871 C; 893 G; 966 T; 0 other;

Query Match             7.1%;  Score 31.6;  DB 13;  Length 3905;
  Best Local Similarity  62.8%;  Pred. No. 2.5;
  Matches    49;  Conservative    0;  Mismatches    29;  Indels    0;  Gaps    0;

Qy      135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
              || ||| ||| ||  ||| ||| |||  ||||| | ||| |||| || ||  ||||||
Db     2049 tcctttaaagaaacctcgcgaaatgtctgggtacacctggatggtcctggagtaacaag 2108

Qy      195 gccaatggcgtcgatgtt 212
              ||||| | | |||||
Db     2109 accaatgaagaagatgtt 2126

RESULT 15
AAC41546
ID   AAC41546 standard; DNA; 946 BP.
XX
AC   AAC41546;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 32273.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000;   2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
```

Mon Sep 10 07:42:21 2001

us-09-540-234-1.rng

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145006.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 03-AUG-1999; 99US-0146388.
PR 04-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 05-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 09-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 10-AUG-1999; 99US-0147493.
PR 11-AUG-1999; 99US-0147935.
PR 12-AUG-1999; 99US-0148171.
PR 13-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 16-AUG-1999; 99US-0148565.
PR 17-AUG-1999; 99US-0148684.
PR 18-AUG-1999; 99US-0149368.
PR 20-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 23-AUG-1999; 99US-0149722.
PR 23-AUG-1999; 99US-0149723.
PR 25-AUG-1999; 99US-0149929.
PR 26-AUG-1999; 99US-0149902.
PR 27-AUG-1999; 99US-0149930.
PR 27-AUG-1999; 99US-0150566.
PR 30-AUG-1999; 99US-0150884.
PR 31-AUG-1999; 99US-0151065.
PR 01-SEP-1999; 99US-0151066.
PR 07-SEP-1999; 99US-0151080.
PR 10-SEP-1999; 99US-0151303.
PR 13-SEP-1999; 99US-0151438.
PR 15-SEP-1999; 99US-0151930.
PR 16-SEP-1999; 99US-0152363.
PR 20-SEP-1999; 99US-0153070.
PR 22-SEP-1999; 99US-0153758.
PR 23-SEP-1999; 99US-0154018.
PR 24-SEP-1999; 99US-0154039.
PR 28-SEP-1999; 99US-0154779.
PR 29-SEP-1999; 99US-0155139.
PR 04-OCT-1999; 99US-0155486.
PR 05-OCT-1999; 99US-0155659.
PR 06-OCT-1999; 99US-0156458.
PR 07-OCT-1999; 99US-0156596.
PR 08-OCT-1999; 99US-0157117.
PR 12-OCT-1999; 99US-0157753.
PR 13-OCT-1999; 99US-0157865.
PR 13-OCT-1999; 99US-0158029.
PR 14-OCT-1999; 99US-0158232.
PR 14-OCT-1999; 99US-0158369.
PR 14-OCT-1999; 99US-0159293.
PR 14-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

```
PR   14-OCT-1999;   99US-0159638.              PR   30-APR-1999;   99US-0132407.
PR   18-OCT-1999;   99US-0159584.              PR   04-MAY-1999;   99US-0132484.
PR   21-OCT-1999;   99US-0160741.              PR   05-MAY-1999;   99US-0132485.
PR   21-OCT-1999;   99US-0160767.              PR   06-MAY-1999;   99US-0132486.
PR   21-OCT-1999;   99US-0160768.              PR   06-MAY-1999;   99US-0132487.
PR   21-OCT-1999;   99US-0160770.              PR   07-MAY-1999;   99US-0132863.
PR   21-OCT-1999;   99US-0160814.              PR   11-MAY-1999;   99US-0134256.
PR   21-OCT-1999;   99US-0160815.              PR   14-MAY-1999;   99US-0134218.
PR   22-OCT-1999;   99US-0160980.              PR   14-MAY-1999;   99US-0134219.
PR   22-OCT-1999;   99US-0160981.              PR   14-MAY-1999;   99US-0134221.
PR   22-OCT-1999;   99US-0160989.              PR   14-MAY-1999;   99US-0134370.
PR   25-OCT-1999;   99US-0161404.              PR   18-MAY-1999;   99US-0134768.
PR   25-OCT-1999;   99US-0161405.              PR   19-MAY-1999;   99US-0134941.
PR   25-OCT-1999;   99US-0161406.              PR   20-MAY-1999;   99US-0135124.
PR   26-OCT-1999;   99US-0161359.              PR   21-MAY-1999;   99US-0135353.
PR   26-OCT-1999;   99US-0161360.              PR   24-MAY-1999;   99US-0135629.
PR   26-OCT-1999;   99US-0161361.              PR   25-MAY-1999;   99US-0136021.
PR   28-OCT-1999;   99US-0161920.              PR   27-MAY-1999;   99US-0136392.
PR   28-OCT-1999;   99US-0161992.              PR   28-MAY-1999;   99US-0136782.
PR   28-OCT-1999;   99US-0161993.              PR   01-JUN-1999;   99US-0137222.
PR   29-OCT-1999;   99US-0162142.              PR   03-JUN-1999;   99US-0137528.
                                               PR   04-JUN-1999;   99US-0137502.
                                               PR   07-JUN-1999;   99US-0137724.
                                               PR   08-JUN-1999;   99US-0138094.
                                               PR   10-JUN-1999;   99US-0138540.
Query Match       7.0%;   Score 31.2;  DB 21;  Length 946;
Best Local Similarity  66.2%;  Pred. No. 1.7;
Matches   45;  Conservative  0;  Mismatches   23;  Indels  0;  Gaps  0;

Qy  205 tcgatgttatcggagccattcgtgtgtttactgtgtagcgcttgatgacttccttgc 264
         |||||  |||||  | ||||  | ||||||||||  ||| ||
Db  472 tcgatccttatcatggcatttgcatggcgttcactgttactgttagctcttcatcctactcttga 531

Qy  265 tgtactcc 272
         ||| | |
Db  532 tgttgtac 539

RESULT 16
AAC41498
ID   AAC41498 standard; DNA; 1278 BP.
XX
AC   AAC41498;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 32103.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   28-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
```

Right column (continued):

```
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   21-JUN-1999;   99US-0139817.
PR   23-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   21-JUL-1999;   99US-0144632.
PR   21-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
```

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match   7.0%;   Score 31.2;   DB 21;   Length 1278;
Best Local Similarity 66.2%;   Pred. No. 2;
Matches 45; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 205 tcgatgttatcggagccaccattcgtcgtgctgttactgtgtagcgcttgatgacttccttgc 264
         ||||| |||| ||||| | |||| || ||||||||||| |||||| ||| ||||
Db 741 tcgatcttatcatggtcattgcatggcgttaactgttagctctcttcatcctactcctagga 800

Qy 265 tgtactcc 272
         ||| | ||
Db 801 tgttgtac 808

RESULT 17
AAC49866
ID   AAC49866 standard; DNA; 1440 BP.
XX
AC   AAC49866;
XX
DT   18-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 62732.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR· 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132863.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134256.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145913.
PR	14-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145918.
PR	18-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137502.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	02-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	06-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	29-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	05-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	06-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	07-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	08-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	12-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	13-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	14-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	14-OCT-1999;	99US-0159637.
PR	21-JUL-1999;	99US-0144884.	PR	14-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	18-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	21-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	21-OCT-1999;	99US-0160770.

```
PR   21-OCT-1999;  99US-0160814.
PR   21-OCT-1999;  99US-0160815.
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   25-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   26-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.

Query Match
Best Local Similarity  66.2%;      Score 31.2;  DB 21;  Length 1440;
Matches  45;  Conservative  0;  Mismatches  23;  Indels  0;  Gaps  0;

Qy  205  tcgatgtatcggagccattcgtgctgtttactgtgtagccgttgatgacttcctgc  264
         ||||||||||| ||||| || |||||||| ||||||||| ||| ||   ||||
Db  741  tcgatcttatcatggtcattgcatggcgtttactgttttagctcttatccctactcttga  800

Qy  265  tgtactcc  272
         ||| | |
Db  801  tgttgtac  808

RESULT 18
AAT39340
ID  AAT39340 standard; DNA; 1868 BP.
XX
AC  AAT39340;
XX
DT  22-JAN-1997  (first entry)
XX
DE  DNA sequence for macrophage receptor with collagenous domain.
XX
KW  Macrophage receptor-collagenous; MARCO; immune defence;
KW  phagocytosis; gene therapy; ss.
XX
OS  Mus sp.
XX
FH  Key              Location/Qualifiers
FT  5'UTR            1..159
FT                   /*tag= a
FT  CDS              160..1716
FT                   /*tag= b
FT  3'UTR            1717..1868
FT                   /*tag= c
FT  polyA_signal     1813..1818
FT                   /*tag= d
XX
PN  WO9626219-A1.
XX
PD  29-AUG-1996.
XX
PF  19-FEB-1996;  96WO-FI00091.
XX
PR  21-FEB-1995;  95US-0392367.
XX
PA  (ELOM/)  ELOMAA O.
PA  (KANG/)  KANGAS M.
PA  (TRYG/)  TRYGGVASON K.
XX
PI  Elomaa O, Kangas M, Tryggvason K;
XX
WPI; 1996-402322/40.
DR  P-PSDB; AAW03561.
XX
PT  New macrophage receptor, MARCO, that binds bacteria but not yeast  -
PT  and related DNA, anti-sense probes and antibodies, involved in
```

```
PT   immune defence and phagocytosis
XX
PS   Claim 1; Fig 2A-B; 62pp; English.
XX
CC   A DNA clone (AAT39340), designated Maf-6, codes for a novel
CC   macrophage receptor having a collagenous structure, named MARCO
CC   (AAW03561), a membrane-bound trimeric protein which probably plays
CC   a role in host defence.  The gene was isolated from a mouse
CC   macrophage cDNA library by screening with human type XIII collagen
CC   DNA.  An isolated clone was used to rescreen the library, yielding
CC   Maf-6.  The DNA or its fragments can be used to produce recombinant
CC   MARCO, as a probe to detect the MARCO gene (or mutations in it that
CC   cause disease), or to correct defects in the MARCO gene.
XX
SQ   Sequence 1868 BP;  524 A;  433 C;  576 G;  335 T;  0 other;

Query Match
Best Local Similarity  51.1%;      Score 29.4;  DB 17;  Length 1868;
Matches  69;  Conservative  0;  Mismatches  66;  Indels  0;  Gaps  0;

Qy    33  ctagtgggtcatctgaaggaaggacaaagacaccccaagccggacacggctactgtag  92
          ||| |||| || |||||||  |||||||||  ||  ||||  |  ||||
Db  1206  caagggtgacacaggaattcaaggacagaaaggcacaaaaggagaatcaggagtcccagg  1265

Qy    93  ctggcacggcataatggaaaacagtccctgccctttgagggtgtctcaaacatgcccc  152
          || |||| ||||  |||||  |||||||  ||||||||||  || ||| |
Db  1266  tcttgtaggcagaaggggagcacactggaagcccctgggctgcagcgtcccaaggagaacc  1325

Qy   153  tcgagacgtcggtga  167
          |||||| ||||||
Db  1326  tggacgagtcggtca  1340

RESULT 19
AAA55966/C
ID  AAA55966 standard; DNA; 21636 BP.
XX
AC  AAA55966;
XX
DT  05-SEP-2000  (first entry)
XX
DE  Human G713 3'-end of intron 2, exon 3 and 3'-regulatory region.
XX
KW  Human; chromosome 13; G713; chromosome 13q31-q33; schizophrenia;
KW  biallelic marker; polymorphism; central nervous disease; detection;
KW  neuroleptic; G713 gene expression inhibitor; genotyping;
KW  brain disorder; psychiatric disorder; bipolar disorder; ds.
XX
OS  Homo sapiens.
XX
PN  WO200022122-A2.
XX
PD  20-APR-2000.
XX
PF  12-OCT-1999;  99WO-IB01730.
XX
PR  13-OCT-1998;  98US-0103955.
PR  30-OCT-1998;  98US-0106457.
XX
PA  (GEST ) GENSET.
XX
PI  Blumenfeld M, Bougueleret L, Chumakov I, Cohen D, Essioux L;
XX
DR  WPI; 2000-317979/27.
XX
PT  Novel polynucleotide of human G713 gene useful for diagnosis and
PT  prophylactic treatment of brain, psychiatric disorders like
PT  schizophrenia and bipolar disorders  -
XX
PS  Claim 1; Page 224-230; 271pp; English.
XX
CC  The present invention describes an isolated, purified or recombinant
```

CC polynucleotide (PN) (I) comprising a contiguous span of 8 to 50
CC nucleotides, where the span includes a G713 or chromosome 13q31-q33
CC related biallelic marker. (I) has neuroleptic activity and can be used
CC as a G713 gene expression inhibitor. (I) can be used genotyping to
CC estimate the frequency of an allele of a G713 or chromosome 13q31-q33
CC related biallelic marker in a population, and of a haplotype for a set
CC of biallelic markers in a population. (I) is also useful in detecting
CC an association between a haplotype and a trait. The frequency is used
CC for detecting an association between a genotype and a trait being
CC schizophrenia. The genotype is used to determine whether an individual
CC is at risk of developing schizophrenia. (I) can also be used as a
CC medicament against several disorders preferably brain, psychiatric
CC disorders such as schizophrenia and bipolar disorder. Early
CC identification of risk of developing schizophrenia is possible, which
CC would enable early and/or prophylactic treatment. AAA55964 to AAA55966
CC represent human G713 genomic DNA sequences; AAA55967 encodes the murine
CC G713 protein AAY90962; AAA55968 encodes the murine G713 protein
CC AAY90963; AAA55992 to AAA56030 represent human chromosome 13q31-q33 locus
CC biallelic markers A12 to A49; AAA55969 to AAA55991, and AAA56031 and
CC AAA56032 represent PCR primers used in the exemplification of the present
CC invention.
XX
SQ Sequence 21636 BP; 7160 A; 3812 C; 4168 G; 6463 T; 33 other;

Query Match   6.6%; Score 29.4; DB 21; Length 21636;
Best Local Similarity 48.5%; Pred. No. 31;
Matches   81; Conservative   0; Mismatches   86; Indels   0; Gaps   0;

Qy   252 atgacttcctgctactccgctggagccattgagtcgttcattgagtcgtgttaccgtgcagcactc 311
             |||||| | || | ||||||||| || |||||| ||| ||||| || |||||
Db   2403 ATTACTTCATCCCTAACTCTGCTTTCAGGGAATTTGGTGAGTTTTCACGTGACATAATT 2344

Qy   312 gaagacttcctgctgtactcgctgcgcagcatagtctgtttgctgcctgtggggaca 371
             ||||| | || || |||| | ||||| || |||| |||||||| |||||
Db   2343 TCACTGAATTTGCTTTTCCCTAGGATCAAGTTCCAAGAGCTCCTTTGTGTTATCCGA 2284

Qy   372 cctgtgtactacagtacttcgttccctttgtagtggtactactct 418
             |||||||| | ||| | ||||||||||| ||||| |||
Db   2283 ATGTTGCTTTTCGAAGCAACTGTTCTTATTTTATGTTACCATGTTT 2237

RESULT 20
AAZ19436/c
ID AAZ19436 standard; cDNA; 1251 BP.
XX
AC AAZ19436;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis antigen 3' Erdsn-2 cDNA sequence.
XX
KW Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
KW immunotherapy; diagnosis; immunisation; vaccine; infection;
KW immune response; skin test; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942076-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999;   99WO-US03268.
XX
PR 05-MAY-1998;   98US-0072967.
PR 18-FEB-1998;   98US-0025197.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527409/44.

XX
PT New antigens from Mycobacterium tuberculosis useful in diagnostic
    skin tests and protective or therapeutic vaccines or compositions
XX
PS Claim 4; Page 256-257; 299pp; English.
XX
CC The present invention describes polypeptides comprising an immunogenic
CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
CC are vaccines and fusion protein containing M. tuberculosis Ag's.
CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
CC other polypeptides fragments, can be used in pharmaceutical compositions
CC or vaccines to generate a protective or therapeutic immune response to
CC M. tuberculosis and as reagents in skin tests for diagnosis of
CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
CC by, T, B or natural killer cells and/or macrophages in
CC tuberculosis-immune subjects. AAZ19249 to AAZ19460 and AAY39083 to
CC AAY39225 are used in the exemplification of the present invention.
XX
SQ Sequence 1251 BP; 190 A; 307 C; 261 G; 215 T; 278 other;

Query Match   6.6%; Score 29.2; DB 20; Length 1251;
Best Local Similarity 37.4%; Pred. No. 9.1;
Matches   55; Conservative   17; Mismatches   75; Indels   0; Gaps   0;

Qy   42 catctgaaggaaggacaaagacacctccaaccggacacggctactgtagctggcacgg 101
            :|| |||||| | ||| : |||| ||| || || :: || ||:
Db   898 SMNCRTWTGGGGGGCSSNAAAACCCCCKAGGGGGTACCCCCMNCNSGGRTMSSTNTG 839

Qy   102 cataatggaaaacagtccctgcctttt gagggtgtcgtctcaaacatgcccctcgagacgt 161
            | : |||| ||| | ||| | ||||||| ||| || ||||| || ||||| |||
Db   838 MCAMASGNGNRGMGGCCCNANSCCNATANAGGGNGNCCSCARRAAATMCCGGAANCCCA 779

Qy   162 cggtgaagccgtagtttgtacagagact 188
            || | |||| ||| |||| ||||
Db   778 MSNTTCGTCCMAASTTTGGCAAAGAWT 752

RESULT 21
AAZ19224/c
ID AAZ19224 standard; cDNA; 1251 BP.
XX
AC AAZ19224;
XX
DT 05-NOV-1999 (first entry)
XX
DE M. tuberculosis recombinant antigen cDNA encoding 3' Erdsn-2.
XX
KW Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity; ss.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9942118-A2.
XX
PD 26-AUG-1999.
XX
PF 17-FEB-1999;   99WO-US03265.
XX
PR 05-MAY-1998;   98US-0072596.
PR 18-FEB-1998;   98US-0024753.
XX
PA (CORI-) CORIXA CORP.
XX
PI Campos-Neto A, Dillon DC, Reed SG, Skeiky YAW, Houghton R;
PI Lodes MJ, Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
XX
DR WPI; 1999-527416/44.
XX
PT New polypeptide comprising antigenic portions of M. tuberculosis
XX
PS Claim 4; Page 302; 323pp; English.
XX

CC   This invention describes novel recombinant antigens and their encoding
CC   nucleic acids derived from Mycobacterium tuberculosis. The novel
CC   polypeptides are useful for detecting M. tuberculosis infection in a
CC   biological sample by detecting antibodies which bind with the
CC   polypeptides, and are useful as vaccines for immunizing against
CC   M. tuberculosis infection. The new detection methods are needed as
CC   current vaccination strategies do not provide 100% immunity.
XX
SQ   Sequence 1251 BP; 190 A; 307 C; 261 G; 215 T; 278 other;

Query Match   6.6%; Score 29.2; DB 20; Length 1251;
Best Local Similarity 37.4%; Pred. No. 9.1;
Matches   55; Conservative   17; Mismatches   75; Indels   0; Gaps   0;

Qy   42 catctgaagaaggacaaagacacctccaagccggacacggctactgtagctggcacgg 101
         :|| |||| || |||| ||| | |||| ||| ||| | ||| :|::
Db 898 SMNCRTWTGGGGGCGSSNAAAAACCCCCKAGGGGGGTACCCCCMNCNSGGRTMSSTNTG 839

Qy 102 cataatggaaaacagtccctgcctttgagggtcgtctcaaacatgcccctgagacgt 161
          :|| :|| ||| |||| || |||| |||||| || :|| :||
Db 838 MCAMASGNGNRGMGGCCCNANSCCNATANAGGGNGNCCSCAARRAAATMCCGGAANCCCA 779

Qy 162 cggtgaagccgtagtttgtacaagact 188
         :: |||| |:||||| |||||
Db 778 MSNTTCGTCCMAASTTGCAAAGAWT 752

RESULT 22
AAX03038
ID   AAX03038 standard; DNA; 3515 BP.
XX
AC   AAX03038;
XX
DT   22-JUN-1999   (first entry)
XX
DE   Human IL-1ra BAC contiguous DNA sequence 83.
XX
KW   Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW   interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW   chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW   growth factors; treatment; IL-1 receptor complex; BAC; ss.
XX
OS   Homo sapiens.
XX
PN   WO990426-A1.
XX
PD   11-FEB-1999.
XX
PF   03-AUG-1998;   98WO-US16102.
XX
PR   02-JUL-1998;   98US-0091650.
PR   04-AUG-1997;   97US-0054646.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   Pan Y;
XX
DR   WPI; 1999-153692/13.
XX
PT   New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT   used to inhibit inflammation and to screen for specific modulators
XX
PS   Example 5; Figure 4; 226pp; English.
XX
CC   AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic
CC   sequences containing alternatively spliced forms of human IL-1ra. Such
CC   fragments are used in the method of the invention which describes the
CC   isolation of a novel human TANGO-77 encoding nucleic acid and protein.
CC   Tango-77 is a member of the cytokine superfamily that is expected to
CC   inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It
CC   may also bind to a new receptor so could regulate other cellular
CC   processes associated with acute or chronic inflammation, e.g. asthma,

CC   chronic myelogenous leukaemia, rheumatoid arthritis, psoriasis and
CC   inflammatory bowel disease. It may also induce or suppress interleukins,
CC   cytokines and growth factors. Modulators of this protein are used to
CC   treat or prevent conditions associated with abnormal levels of
CC   inflammation, or activity of IL-1 or its receptor complex.
XX
SQ   Sequence 3515 BP; 929 A; 724 C; 764 G; 1094 T; 4 other;

Query Match   6.6%; Score 29.2; DB 20; Length 3515;
Best Local Similarity 52.5%; Pred. No. 15;
Matches   64; Conservative   0; Mismatches   58; Indels   0; Gaps   0;

Qy 109 gaaaacagtcctgccctttgagggtcgtctcaaacatgcccctcgagacgtcggtgaa 168
         ||| ||||| | ||||| |||| ||||||||||||||||| |||||||| ||
Db 1163 gacagtagtcccctcatctgaggaggggcgtgttccaagcccctcagtgaatgcctgaa 1222

Qy 169 gccgtagtttgtacaagactaacaacgccaatggcgtcgatgtgtatcggagccattcgtg 228
         ||||| ||| |||| ||| || || |||| || ||| |||| |||||
Db 1223 actgtggatagtaccaactctatgtctatgatttcctataattaatacatgcctg 1282

Qy 229 tg 230
         ||
Db 1283 tg 1284

RESULT 23
AAX02998/c
ID   AAX02998 standard; DNA; 11901 BP.
XX
AC   AAX02998;
XX
DT   22-JUN-1999   (first entry)
XX
DE   Human IL-1ra BAC contiguous DNA sequence 43.
XX
KW   Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW   interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW   chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW   growth factors; treatment; IL-1 receptor complex; BAC; ss.
XX
OS   Homo sapiens.
XX
PN   WO990426-A1.
XX
PD   11-FEB-1999.
XX
PF   03-AUG-1998;   98WO-US16102.
XX
PR   02-JUL-1998;   98US-0091650.
PR   04-AUG-1997;   97US-0054646.
XX
PA   (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX
PI   Pan Y;
XX
DR   WPI; 1999-153692/13.
XX
PT   New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT   used to inhibit inflammation and to screen for specific modulators
XX
PS   Example 5; Figure 3; 226pp; English.
XX
CC   AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic
CC   sequences containing alternatively spliced forms of human IL-1ra. Such
CC   fragments are used in the method of the invention which describes the
CC   isolation of a novel human TANGO-77 encoding nucleic acid and protein.
CC   Tango-77 is a member of the cytokine superfamily that is expected to
CC   inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It
CC   may also bind to a new receptor so could regulate other cellular
CC   processes associated with acute or chronic inflammation, e.g. asthma,
CC   chronic myelogenous leukaemia, rheumatoid arthritis, psoriasis and
CC   inflammatory bowel disease. It may also induce or suppress interleukins,

CC   cytokines and growth factors. Modulators of this protein are used to
CC   treat or prevent conditions associated with abnormal levels of
CC   inflammation, or activity of IL-1 or its receptor complex.
XX
SQ   Sequence 11901 BP; 3552 A; 2724 C; 2438 G; 3179 T; 8 other;

Query Match
Best Local Similarity   6.6%; Score 29.2; DB 20; Length 11901;
Matches   64; Conservative   0; Mismatches   58; Indels   0; Gaps   0;

Qy   109 gaaaacagtcctgccctttgaggtcgtctcaaacatgcccctcgagacgtcggtgaa 168
             ||||||| | |||||||||| || |||| |||| | |||||||| |||
Db   4977 GACAGTAGTCCCCCTCCATCTGAGAGGGGCGTGTTCCAAGCCCCTCAGTGAATGCCTGAA 4918
Qy   169 gccgtagtttgtacaagactaacaacgccaatggcgtcgatgttatcggagccattcgtg 228
             | ||| | ||| || | ||| || |||| ||| || || ||| |||
Db   4917 ACTGTGGATAGTACCCAACTCTATATGTCTATGATTTTCCTATAAATTAATACATGCCTG 4858
Qy   229 tg 230
             ||
Db   4857 TG 4856

RESULT 24
AAC26317
ID   AAC26317 standard; cDNA; 411 BP.
XX
AC   AAC26317;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 30392.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999; 99US-0122487.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 30392; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.

XX
SQ   Sequence 411 BP; 138 A; 69 C; 84 G; 119 T; 1 other;

Query Match
Best Local Similarity 54.7%; Score 29; DB 21; Length 411;
Matches   76; Conservative   1; Mismatches   61; Indels   1; Gaps   1;

Qy   40 gtcatctgaaggaaggacaaagacacctccaagccggacaacggctactgtagctggca- 98
             ||| || ||||||| || | |||| | || ||||| ||| |||| ||||
Db   84 gttatcatacagaaataaaatagacatctatcagcttgacactgacactagtgcac 143
Qy   99 cggcatatgaaaacagtcccgcccctttgagggtcgtctcaaacatgcccctcgaga 158
             | || | |||||||| | || ||||| | || |||||||||||| ||||
Db   144 catcacaaatgamaagagaattccctcgttaattctttcagagtctgtccttgcagt 203
Qy   159 cgtcggtgaagccgtagtt 177
             || | ||| |||
Db   204 ggctggagatgaagaagtt 222

RESULT 25
AAC06849
ID   AAC06849 standard; cDNA; 488 BP.
XX
AC   AAC06849;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 10924.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999; 99US-0122487.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 10924; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 488 BP; 169 A; 90 C; 113 G; 116 T; 0 other;

```
Query Match      6.5%;  Score 29;  DB 21;  Length 488;
Best Local Similarity  54.1%;  Pred. No. 6.8;
Matches  59;  Conservative  0;  Mismatches  50;  Indels  0;  Gaps  0;

Qy    1 gggtccacggccccttatgtatgcaggagcgcctagtggtcatctgaaggaaaggacaa 60
          || ||| ||| | || | | | | ||   |  | | | || | | || |||| |||
Db  130 gagtcaaaggacctgtaggaagctaaagtctgtggaagctattctagaagaaagcactg 189

Qy   61 agacacctccaagccgacacggctactgtagctgcacggcataatgg 109
          | | || ||||   ||  || ||| || ||  ||  |||||||
Db  190 aaaaactcaaagcttgtcactgcgcaacagcaggatggagataatgg 238
```

```
RESULT 26
AAC77597
ID   AAC77597 standard; cDNA; 676 BP.
XX
AC   AAC77597;
XX
DT   08-FEB-2001   (first entry)
XX
DE   Human ORFX ORF3152 polynucleotide sequence SEQ ID NO:6303.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO20005847-3-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000;   2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB43388.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 5488-5489; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 676 BP; 223 A; 130 C; 146 G; 177 T; 0 other;
```

```
Query Match      6.5%;  Score 29;  DB 21;  Length 676;
Best Local Similarity  54.1%;  Pred. No. 7.9;
Matches  59;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 gggtccacggccccttatgtatgcaggagcgcctagtggtcatctgaaggaaaggacaa 60
          || ||| ||| | || | | | | ||   |  | | | || | | || |||| |||
Db  425 gagtcaaaggacctgtaggaagctaaagtctgtggaagctattctagaagaaagcactg 484

Qy   61 agacacctccaagccgacacggctactgtagctgcacggcataatgg 109
          | | || ||||   ||  || ||| || ||  ||  |||||||
Db  485 aaaaactcaaagcttgtcactgcgcaacagcaggatggagataatgg 533
```

```
RESULT 27
AAC52561
ID   AAC52561 standard; DNA; 1193 BP.
XX
AC   AAC52561;
XX
DT   18-OCT-2000   (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 71747.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132408.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
```

```
PR  07-MAY-1999;  99US-0132863.
PR  11-MAY-1999;  99US-0133256.
PR  14-MAY-1999;  99US-0134218.
PR  14-MAY-1999;  99US-0134219.
PR  14-MAY-1999;  99US-0134221.
PR  18-MAY-1999;  99US-0134370.
PR  18-MAY-1999;  99US-0134768.
PR  19-MAY-1999;  99US-0134941.
PR  20-MAY-1999;  99US-0135124.
PR  21-MAY-1999;  99US-0135353.
PR  24-MAY-1999;  99US-0135629.
PR  25-MAY-1999;  99US-0136021.
PR  27-MAY-1999;  99US-0136392.
PR  28-MAY-1999;  99US-0136782.
PR  01-JUN-1999;  99US-0137222.
PR  03-JUN-1999;  99US-0137528.
PR  04-JUN-1999;  99US-0137502.
PR  07-JUN-1999;  99US-0137724.
PR  08-JUN-1999;  99US-0138094.
PR  10-JUN-1999;  99US-0138540.
PR  10-JUN-1999;  99US-0138847.
PR  14-JUN-1999;  99US-0139119.
PR  16-JUN-1999;  99US-0139452.
PR  16-JUN-1999;  99US-0139453.
PR  17-JUN-1999;  99US-0139492.
PR  18-JUN-1999;  99US-0139454.
PR  18-JUN-1999;  99US-0139455.
PR  18-JUN-1999;  99US-0139456.
PR  18-JUN-1999;  99US-0139457.
PR  18-JUN-1999;  99US-0139458.
PR  18-JUN-1999;  99US-0139459.
PR  18-JUN-1999;  99US-0139460.
PR  18-JUN-1999;  99US-0139461.
PR  18-JUN-1999;  99US-0139462.
PR  18-JUN-1999;  99US-0139463.
PR  18-JUN-1999;  99US-0139750.
PR  18-JUN-1999;  99US-0139763.
PR  21-JUN-1999;  99US-0139817.
PR  22-JUN-1999;  99US-0139899.
PR  23-JUN-1999;  99US-0140353.
PR  23-JUN-1999;  99US-0140354.
PR  24-JUN-1999;  99US-0140695.
PR  28-JUN-1999;  99US-0140823.
PR  29-JUN-1999;  99US-0140991.
PR  30-JUN-1999;  99US-0141287.
PR  01-JUL-1999;  99US-0141842.
PR  01-JUL-1999;  99US-0142154.
PR  02-JUL-1999;  99US-0142055.
PR  06-JUL-1999;  99US-0142390.
PR  08-JUL-1999;  99US-0142803.
PR  09-JUL-1999;  99US-0142920.
PR  12-JUL-1999;  99US-0142977.
PR  13-JUL-1999;  99US-0143542.
PR  14-JUL-1999;  99US-0143624.
PR  15-JUL-1999;  99US-0144005.
PR  16-JUL-1999;  99US-0144085.
PR  16-JUL-1999;  99US-0144086.
PR  19-JUL-1999;  99US-0144325.
PR  19-JUL-1999;  99US-0144331.
PR  19-JUL-1999;  99US-0144332.
PR  19-JUL-1999;  99US-0144333.
PR  19-JUL-1999;  99US-0144334.
PR  19-JUL-1999;  99US-0144335.
PR  20-JUL-1999;  99US-0144352.
PR  20-JUL-1999;  99US-0144632.
PR  21-JUL-1999;  99US-0144884.
PR  21-JUL-1999;  99US-0144814.
PR  21-JUL-1999;  99US-0145086.
PR  22-JUL-1999;  99US-0145088.
PR  22-JUL-1999;  99US-0145085.
PR  22-JUL-1999;  99US-0145087.
PR  22-JUL-1999;  99US-0145089.
PR  22-JUL-1999;  99US-0145192.

PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  03-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  23-AUG-1999;  99US-0149930.
PR  25-AUG-1999;  99US-0150566.
PR  26-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  27-AUG-1999;  99US-0151066.
PR  27-AUG-1999;  99US-0151080.
PR  30-AUG-1999;  99US-0151303.
PR  31-AUG-1999;  99US-0151438.
PR  01-SEP-1999;  99US-0151930.
PR  07-SEP-1999;  99US-0152363.
PR  10-SEP-1999;  99US-0153070.
PR  13-SEP-1999;  99US-0153758.
PR  15-SEP-1999;  99US-0154018.
PR  16-SEP-1999;  99US-0154039.
PR  20-SEP-1999;  99US-0154779.
PR  22-SEP-1999;  99US-0155139.
PR  23-SEP-1999;  99US-0155486.
PR  24-SEP-1999;  99US-0155659.
PR  28-SEP-1999;  99US-0156458.
PR  29-SEP-1999;  99US-0156596.
PR  04-OCT-1999;  99US-0157117.
PR  05-OCT-1999;  99US-0157753.
PR  06-OCT-1999;  99US-0157865.
PR  07-OCT-1999;  99US-0158029.
PR  08-OCT-1999;  99US-0158232.
PR  12-OCT-1999;  99US-0158369.
PR  13-OCT-1999;  99US-0159293.
PR  13-OCT-1999;  99US-0159294.
PR  13-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  14-OCT-1999;  99US-0159638.
PR  18-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
PR  21-OCT-1999;  99US-0160815.
```

```
PR   22-OCT-1999;  99US-0160980.
PR   22-OCT-1999;  99US-0160981.
PR   22-OCT-1999;  99US-0160989.
PR   25-OCT-1999;  99US-0161404.
PR   25-OCT-1999;  99US-0161405.
PR   25-OCT-1999;  99US-0161406.
PR   25-OCT-1999;  99US-0161359.
PR   26-OCT-1999;  99US-0161360.
PR   26-OCT-1999;  99US-0161361.
PR   28-OCT-1999;  99US-0161920.
PR   28-OCT-1999;  99US-0161992.
PR   28-OCT-1999;  99US-0161993.
PR   29-OCT-1999;  99US-0162142.
XX
```

```
Query Match             6.5%;    Score 29; DB 21; Length 1193;
Best Local Similarity  63.8%;    Pred. No. 10;
Matches    44; Conservative    0; Mismatches    25; Indels   0; Gaps   0;

Qy  287 gagtgctgttaccgtgcagcactcgaagacttcctgctactcgctggcagcatagt 346
        || |||| ||| || |||||||| ||  ||||||| ||| |||||
Db  259 gaaggctgctgcctttcccagtcgaagaaatccgaactgcttaatggtagcatatg 318

Qy  347 ctgtttgct 355
        ||| |||||
Db  319 cggtatgct 327
```

RESULT 28
AAA13984
```
ID   AAA13984 standard; cDNA; 1889 BP.
XX
AC   AAA13984;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Taxus canadensis geranylgeranyl diphosphate synthase encoding cDNA.
XX
KW   Geranylgeranyl diphosphate synthase; GGPP synthase; yew; cytostatic;
KW   anticancer; Taxus; diterpene; paclitaxel; identification; plant;
KW   Taxomyces andreanae; Penicillium raistrickii; microorganism; ss.
XX
OS   Taxus canadensis.
XX
FH   Key           Location/Qualifiers
FT   CDS           308..1489
FT                 /*tag= a
FT                 /product= "geranylgeranyl diphosphate synthase"
XX
PN   US6043072-A.
XX
PD   28-MAR-2000.
XX
PF   05-NOV-1998;  98US-0187050.
XX
PR   05-NOV-1998;  98US-0187050.
XX
PA   (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
PI   Croteau RB,  Hefner JL;
XX
DR   WPI;  2000-282526/24.
DR   P-PSDB; AAY82651.
XX
PT   Nucleic acid encoding geranylgeranyl diphosphate is useful for
PT   producing paclitaxel and other diterpenes that are useful as anticancer
PT   drugs -
XX
PS   Claim 4; Column 33-36; 57pp; English.
XX
CC   The present sequence encodes a geranylgeranyl diphosphate (GGPP)
CC   synthase protein. The GGPP synthase has cytostatic activity. A vector
CC   encoding GGPP synthase is useful in increasing GGPP synthase levels in
CC   a host cell preferably Taxus (Yew) cell and thereby facilitates
CC   production, isolation and purification of larger amounts of GGPP
CC   synthase in plants. GGPP synthase is useful in obtaining expression or
CC   enhanced expression of GGPP and other diterpenes, such as paclitaxel,
CC   useful as anticancer drugs. Isolated nucleic acids encoding GGPP
CC   synthase or hybridising with GGPP synthase encoding nucleic acids are
CC   used for identifying genes encoding GGPP synthase from microorganisms
CC   such as Taxomyces andreanae and Penicillium raistrickii.
XX
SQ   Sequence 1889 BP; 535 A; 300 C; 469 G; 585 T; 0 other;
```

```
Query Match             6.5%;    Score 29; DB 21; Length 1889;
Best Local Similarity  50.4%;    Pred. No. 13;
Matches    71; Conservative    0; Mismatches    70; Indels   0; Gaps   0;

Qy  278 gagccattgagtgctgtttaccgtgcagcactcgaagacttcctgctactcgctgg 337
        |||||||  || ||||||  || ||||  ||| |||  ||||   |||  |||| ||
Db 1401 gagccaaggaagagctgtcatccttgatcagataaaggctgcacctttgggtcttg 1460

Qy  338 cagcatagtctgtttgctgctgcctggtctgggacacctgtactacagatacttcc 397
        || |||| || |||||  || ||| |||| ||  ||| ||| ||  |||| |||||
Db 1461 cagattacattgcattcaggcagacaaactgagaacaaagctgaaagctattcttacacatatc 1520

Qy  398 ctttgtagtggtactactct 418
        ||| |||  ||  || ||||
Db 1521 atctgttttttgacatct 1541
```

RESULT 29
AAC59139
```
ID   AAC59139 standard; cDNA; 2812 BP.
XX
AC   AAC59139;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Human secreted protein coding sequence SEQ ID NO: 42.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055177-A2.
XX
PD   21-SEP-2000.
XX
PF   09-MAR-2000;  2000WO-US06058.
XX
PR   12-MAR-1999;  99US-0124145.
PR   03-DEC-1999;  99US-0168654.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-638177/61.
DR   P-PSDB; AAB28043.
XX
PT   Novel nucleic acids encoding 49 human secreted proteins useful for
PT   treating cancers, hyperproliferative disorders, inflammatory disorders,
PT   neurological disorders and cardiovascular disorders -
XX
PS   Claim 1; Page 335-336; 389pp; English.
XX
CC   The invention relates to the isolation of genes AAA59108-A59156 encoding
CC   49 human secreted proteins AAB28012-B28060. The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (SEQID1) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
```

```
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 2812 BP; 902 A; 482 C; 638 G; 777 T; 13 other;

Query Match             6.5%;  Score 29;  DB 21;  Length 2812;
Best Local Similarity   54.1%;  Pred. No. 16;
Matches   59;  Conservative   0;  Mismatches   50;  Indels   0;  Gaps   0;

Qy      1 gggtccacgggccccttatgtgatgcaggagccgcctagtgggtcatctgaaggaaaggacaa 60
            ||  |||||||   |||||| | |  ||||    |||  ||||||  ||| ||||| |  ||
Db    425 gagtcaaggaccctgtaggaagcctaaagtctgtggaagctattcctagagagaaagccactg 484

Qy     61 agacacctccaagccggacacggctactgtagctgctgcacggcataatgg 109
            |||||| |||| || ||| |  ||||  ||||| | |   |||||||
Db    485 aaaaactcaaaagcttgtcactgcgacaacgcaggatgggagataatgg 533


RESULT 30
AAC36342/C
ID   AAC36342 standard; DNA; 1889 BP.
XX
AC   AAC36342;
XX
DT   17-OCT-2000 (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 13448.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;  99US-0121825.
PR   05-MAR-1999;  99US-0123180.
PR   09-MAR-1999;  99US-0123548.
PR   23-MAR-1999;  99US-0125788.
PR   25-MAR-1999;  99US-0126264.
PR   29-MAR-1999;  99US-0126785.
PR   01-APR-1999;  99US-0127462.
PR   06-APR-1999;  99US-0128234.
PR   08-APR-1999;  99US-0128714.
PR   16-APR-1999;  99US-0129845.
PR   19-APR-1999;  99US-0130077.
PR   21-APR-1999;  99US-0130449.
PR   23-APR-1999;  99US-0130510.
PR   28-APR-1999;  99US-0130891.
PR   30-APR-1999;  99US-0131449.
PR   30-APR-1999;  99US-0132048.
PR   30-APR-1999;  99US-0132407.
PR   04-MAY-1999;  99US-0132484.
PR   05-MAY-1999;  99US-0132485.
PR   06-MAY-1999;  99US-0132486.
PR   06-MAY-1999;  99US-0132487.
PR   07-MAY-1999;  99US-0132863.
PR   11-MAY-1999;  99US-0134256.
PR   14-MAY-1999;  99US-0134218.
PR   14-MAY-1999;  99US-0134219.
PR   14-MAY-1999;  99US-0134221.
PR   14-MAY-1999;  99US-0134370.
PR   19-MAY-1999;  99US-0134768.
PR   19-MAY-1999;  99US-0134941.
PR   20-MAY-1999;  99US-0135124.
PR   21-MAY-1999;  99US-0135353.
PR   24-MAY-1999;  99US-0135629.
PR   25-MAY-1999;  99US-0136021.
PR   27-MAY-1999;  99US-0136392.
PR   28-MAY-1999;  99US-0136782.
PR   01-JUN-1999;  99US-0137222.
PR   03-JUN-1999;  99US-0137528.
PR   04-JUN-1999;  99US-0137502.
PR   07-JUN-1999;  99US-0137724.
PR   08-JUN-1999;  99US-0138094.
PR   10-JUN-1999;  99US-0138540.
PR   10-JUN-1999;  99US-0138847.
PR   14-JUN-1999;  99US-0139119.
PR   16-JUN-1999;  99US-0139452.
PR   16-JUN-1999;  99US-0139453.
PR   17-JUN-1999;  99US-0139492.
PR   18-JUN-1999;  99US-0139454.
PR   18-JUN-1999;  99US-0139455.
PR   18-JUN-1999;  99US-0139456.
PR   18-JUN-1999;  99US-0139457.
PR   18-JUN-1999;  99US-0139458.
PR   18-JUN-1999;  99US-0139459.
PR   18-JUN-1999;  99US-0139460.
PR   18-JUN-1999;  99US-0139461.
PR   18-JUN-1999;  99US-0139462.
PR   18-JUN-1999;  99US-0139463.
PR   18-JUN-1999;  99US-0139750.
PR   21-JUN-1999;  99US-0139763.
PR   22-JUN-1999;  99US-0139817.
PR   22-JUN-1999;  99US-0139899.
PR   23-JUN-1999;  99US-0140353.
PR   23-JUN-1999;  99US-0140354.
PR   24-JUN-1999;  99US-0140695.
PR   28-JUN-1999;  99US-0140823.
PR   29-JUN-1999;  99US-0140991.
PR   30-JUN-1999;  99US-0141287.
PR   01-JUL-1999;  99US-0141842.
PR   01-JUL-1999;  99US-0142154.
PR   02-JUL-1999;  99US-0142055.
PR   06-JUL-1999;  99US-0142390.
PR   08-JUL-1999;  99US-0142803.
PR   09-JUL-1999;  99US-0142920.
PR   12-JUL-1999;  99US-0142977.
PR   13-JUL-1999;  99US-0143542.
PR   14-JUL-1999;  99US-0143624.
PR   15-JUL-1999;  99US-0144005.
PR   16-JUL-1999;  99US-0144085.
PR   16-JUL-1999;  99US-0144086.
PR   19-JUL-1999;  99US-0144325.
PR   19-JUL-1999;  99US-0144331.
PR   19-JUL-1999;  99US-0144332.
PR   19-JUL-1999;  99US-0144333.
PR   19-JUL-1999;  99US-0144334.
PR   19-JUL-1999;  99US-0144335.
PR   20-JUL-1999;  99US-0144352.
PR   20-JUL-1999;  99US-0144632.
PR   20-JUL-1999;  99US-0144884.
PR   21-JUL-1999;  99US-0144814.
PR   21-JUL-1999;  99US-0145086.
PR   21-JUL-1999;  99US-0145088.
PR   22-JUL-1999;  99US-0145085.
PR   22-JUL-1999;  99US-0145087.
PR   22-JUL-1999;  99US-0145089.
```

us-09-540-234-1.rng

```
PR  21-OCT-1999;  99US-0160815.
PR  22-OCT-1999;  99US-0160980.
PR  22-OCT-1999;  99US-0160981.
PR  22-OCT-1999;  99US-0160989.
PR  25-OCT-1999;  99US-0161404.
PR  25-OCT-1999;  99US-0161405.
PR  25-OCT-1999;  99US-0161406.
PR  26-OCT-1999;  99US-0161359.
PR  26-OCT-1999;  99US-0161360.
PR  26-OCT-1999;  99US-0161361.
PR  28-OCT-1999;  99US-0161920.
PR  28-OCT-1999;  99US-0161992.
PR  28-OCT-1999;  99US-0161993.
PR  29-OCT-1999;  99US-0162142.


Query Match                6.5%;     Score 28.8;  DB 21;    Length 1889;
Best Local Similarity   48.8%;    Pred. No. 15;
Matches   78;  Conservative    0;  Mismatches  82;  Indels    0;  Gaps    0;

Qy         45 ctgaggaaggacaaagacacctccaagccgacacggctactgtagctggcacggcat 104
              |||||||   |||| | |||||||  | ||| |||| |||  ||| || ||||
Db        386 CTGCAGATGCGGCGGGAAGAAACCGAGATGCAGACGAGGTAGCACGCCGCCGCCG 327

Qy        105 aatgaaaacagtcctgccgcctttgagggtcgtctcaaacatgccctcgagacgtcgg 164
              |||||||  |||| ||| || ||||   || ||| ||| || | | |||||  | |||
Db        326 AAGGTAACGCGGAGGATGGGTTGTTGGACCACTTAGCCGCCACGATCTGTGCAACACCGA 267

Qy        165 tgaagccgtagtttgtacaagactaacaacgccaatggcg 204
              ||  |  | ||||| |||  ||| | | | ||  ||| |||
Db        266 TGTTGCTACAGTTCTACGGGCTTAATGCTGTCAGCTGACG 227
```

RESULT 31
```
AAA10594/c
ID   AAA10594 standard; DNA; 10732 BP.
XX
AC   AAA10594;
XX
DT   29-JUN-2000  (first entry)
XX
DE   Gene encoding a subunit of cellulose synthase.
XX
KW   Cellulose synthase; cellulose production; increase yield; ds.
XX
OS   Vigna angularis.
XX
XX
PN   JP2000060568-A.
XX
PD   29-FEB-2000.
XX
PF   26-AUG-1998;  98JP-0239998.
XX
PR   26-AUG-1998;  98JP-0239998.
XX
PA   (MIZU/) MIZUNO K.
PA   (OJIP ) OJI PAPER CO.
XX
DR   WPI; 2000-342371/30.
DR   P-PSDB; AAY85179.
XX
PT   A gene encoding a cellulose synthetic equipment - for the improvement
PT   in the amount of cellulose synthesised in a plant body
XX
PS   Claim 2; Page 14-21; 32pp; Japanese.
XX
CC   This sequence represents a gene encoding a subunit of the cellulose
CC   synthase complex of Vigna angularis. The invention relates to subunits of
CC   cellulose synthetic equipment, that can be used to increase the amount of
CC   cellulose synthesised by a plant. The proteins and genes encoding them
CC   can also be used to improve the properties of the cellulose being
CC   produced by a plant.
XX
SQ   Sequence 10732 BP; 3149 A; 1212 C; 2074 G; 2046 T; 2251 other;
```

```
PR  22-JUL-1999;  99US-0145192.
PR  23-JUL-1999;  99US-0145145.
PR  23-JUL-1999;  99US-0145218.
PR  23-JUL-1999;  99US-0145224.
PR  26-JUL-1999;  99US-0145276.
PR  27-JUL-1999;  99US-0145913.
PR  27-JUL-1999;  99US-0145918.
PR  27-JUL-1999;  99US-0145919.
PR  28-JUL-1999;  99US-0145951.
PR  02-AUG-1999;  99US-0146386.
PR  02-AUG-1999;  99US-0146388.
PR  02-AUG-1999;  99US-0146389.
PR  03-AUG-1999;  99US-0147038.
PR  04-AUG-1999;  99US-0147204.
PR  04-AUG-1999;  99US-0147302.
PR  05-AUG-1999;  99US-0147192.
PR  05-AUG-1999;  99US-0147260.
PR  06-AUG-1999;  99US-0147303.
PR  06-AUG-1999;  99US-0147416.
PR  09-AUG-1999;  99US-0147493.
PR  09-AUG-1999;  99US-0147935.
PR  10-AUG-1999;  99US-0148171.
PR  11-AUG-1999;  99US-0148319.
PR  12-AUG-1999;  99US-0148341.
PR  13-AUG-1999;  99US-0148565.
PR  13-AUG-1999;  99US-0148684.
PR  16-AUG-1999;  99US-0149368.
PR  17-AUG-1999;  99US-0149175.
PR  18-AUG-1999;  99US-0149426.
PR  20-AUG-1999;  99US-0149722.
PR  20-AUG-1999;  99US-0149723.
PR  20-AUG-1999;  99US-0149929.
PR  23-AUG-1999;  99US-0149902.
PR  25-AUG-1999;  99US-0149930.
PR  26-AUG-1999;  99US-0150566.
PR  27-AUG-1999;  99US-0150884.
PR  27-AUG-1999;  99US-0151065.
PR  30-AUG-1999;  99US-0151066.
PR  31-AUG-1999;  99US-0151080.
PR  01-SEP-1999;  99US-0151303.
PR  07-SEP-1999;  99US-0151438.
PR  10-SEP-1999;  99US-0151930.
PR  13-SEP-1999;  99US-0152363.
PR  15-SEP-1999;  99US-0153070.
PR  16-SEP-1999;  99US-0153758.
PR  20-SEP-1999;  99US-0154018.
PR  22-SEP-1999;  99US-0154039.
PR  03-SEP-1999;  99US-0154779.
PR  24-SEP-1999;  99US-0155139.
PR  28-SEP-1999;  99US-0155486.
PR  29-SEP-1999;  99US-0155659.
PR  04-OCT-1999;  99US-0156458.
PR  05-OCT-1999;  99US-0156596.
PR  06-OCT-1999;  99US-0157117.
PR  07-OCT-1999;  99US-0157753.
PR  08-OCT-1999;  99US-0157865.
PR  12-OCT-1999;  99US-0158029.
PR  13-OCT-1999;  99US-0158232.
PR  13-OCT-1999;  99US-0158369.
PR  14-OCT-1999;  99US-0159293.
PR  14-OCT-1999;  99US-0159294.
PR  14-OCT-1999;  99US-0159295.
PR  14-OCT-1999;  99US-0159329.
PR  14-OCT-1999;  99US-0159330.
PR  14-OCT-1999;  99US-0159331.
PR  14-OCT-1999;  99US-0159637.
PR  18-OCT-1999;  99US-0159638.
PR  21-OCT-1999;  99US-0159584.
PR  21-OCT-1999;  99US-0160741.
PR  21-OCT-1999;  99US-0160767.
PR  21-OCT-1999;  99US-0160768.
PR  21-OCT-1999;  99US-0160770.
PR  21-OCT-1999;  99US-0160814.
```

Query Match   6.5%;   Score 28.8; DB 21; Length 10732;
Best Local Similarity 16.08%; Pred. No. 35;
Matches   38; Conservative 94; Mismatches 106; Indels   0; Gaps   0;

Qy   203 cgtcgatgttatcggagccattcgtgtcgtgttactgtgtagcgcttgatgacttcctt 262
           ::: :: ||: :: :: :|: :| ||: || : :|:: ||: |: || :
Db 10060 YRASRGTBSDDCCTBSRYSCYSYDASRYANCDTBCYTTBAKYRARCYDAYAKRCNSTSRY 10001

Qy   263 gctgtactccgctggagccattgagtgcgtgttaccgtgcagccactcgaagacttcct 322
           : :::: :|:: ||:| :|:|::: :||| ||||| | || : |: : |:
Db 10000 RAYSDSTCYTRCRCSRNCYSTYSYSYTYRASTBTTYCYTCBCSRRCYSRSYSSTCNCY 9941

Qy   323 tgctgtactcgctggcagcatagtgttgctgctgctgggacacctgtacta 382
           : ::: :| ||: :|:::   :|:||| |:|||::|::|:::::|:|||
Db   9940 SYCCYTSRYSSTTTNTSRCTTYSYNSTTBYBYSYSSTYSSRGYSSRGDSRGNCYYNST 9881

Qy   383 cagtacttcgtttccccttgtagtggtactactccttagccatgtactcgtgtctgtt 440
           : :::| :: ||::|:|:|||:|| :| : ||:: : ||| | | :|:|:||
Db   9880 NCYDASTSDTBYSRCCYTYSYSSTDSTDSTNSTTBSDCYTTTBSRSTSDSTSTY 9823

RESULT 32
AAF74196/c
ID   AAF74196 standard; DNA; 647 BP.
XX
AC   AAF74196;
XX
DT   02-MAY-2001 (first entry)
XX
DE   DNA encoding evironmental stress tolerant protein SEQ ID 19.
XX
KW   Environmental stress resistance; salt; heat; desert; transgenic plant;
KW   ds.
XX
OS   Mesembryanthemum crystallinum.
XX
PN   WO200106006-A1.
XX
PD   25-JAN-2001.
XX
PF   19-JUL-2000; 2000WO-JP04862.
XX
PR   19-JUL-1999; 99JP-0235910.
PR   24-MAR-2000; 2000JP-0085377.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Yamada A, Ozeki Y, Saito T;
XX
DR   WPI; 2001-147355/15.
DR   P-PSDB; AAB06617.
XX
PT   Screening method to obtain DNA encoding environmental stress resistance
PT   factor, useful for producing transgenic plants resistant to
PT   environmental stress -
XX
PS   Claim 41; Page 101-102; 167pp; Japanese.
XX
CC   Polynucleotide sequences AAF74187 - AAF74218 encode proteins
CC   AAB80608 - AAB80639, which impart environmental stress resistance. The
CC   invention relates to a method for identifying DNA encoding proteins
CC   imparting environmental stress resistance. The method comprises inserting
CC   cDNA from a library originating in a salt-resistant organism into a host
CC   cell, culturing the transformants under conditions in which the
CC   untransformed host does not grow well, and selecting for viable clones.
CC   The method is useful for obtaining DNA encoding environmental stress
CC   resistance factors. The DNA encoding proteins conferring environmental
CC   stress resistance, can be used in the production of plants resistant to
CC   environmental stress, which can be cultivated in unfavourable
CC   environments such as deserts, salt damaged ground, cold regions and the
CC   oceans. They can be used for increasing the area of land covered by green

CC   plants, and desert greening and afforestation, in order to counter the
CC   effects of the increase in atmospheric carbon dioxide concentration. PCR
CC   primers AAF74219 and AAF74220 are used in an example illustrating the
CC   method of the invention.
XX
SQ   Sequence 647 BP; 180 A; 133 C; 150 G; 184 T; 0 other;

Query Match   6.48%;   Score 28.6; DB 22; Length 647;
Best Local Similarity 52.08%; Pred. No. 11;
Matches   64; Conservative   0; Mismatches 59; Indels   0; Gaps   0;

Qy   241 tgtagcgcttgatgacttcctgctgtactccgcttggagccattggagtgctgttacc 300
           || ::|:| :: ||:|| || || |:| | || |||| | : | : | ||| | |:
Db   346 TGAAGAGGTCAATGACTCTCTTGTGAACACGCAACTCCAAATCGTCAAAGGTGTTGGTTC 287

Qy   301 gtgcagcactcgaagacttcctgctgctgctgctgcagcatagtcgtgttctgctgcctg 360
           | |:| :| ||:|:: |:|| |:|: |:| ||| |:| ||| :| |:| || | |||
Db   286 CTTCACCACAGGGAGACTTCCTTGTCGTGATCTTCAGAACCTTCGGTGGGCATCCTCACTG 227

Qy   361 gtc 363
           |||
Db   226 GTC 224

RESULT 33
AAC83722/c
ID   AAC83722 standard; cDNA; 5113 BP.
XX
AC   AAC83722;
XX
DT   02-MAR-2001 (first entry)
XX
DE   Rat laminin 5 cDNA, SEQ ID NO: 11.
XX
KW   Rat; laminin 5; vulnerary; antiulcer; antiinflammatory; antidiabetic;
KW   cell adhesion promoter; wound healing; ulcers; burn; skin graft;
KW   periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.
XX
OS   Rattus norvegicus.
XX
PN   WO200066731-A2.
XX
PD   09-NOV-2000.
XX
PF   28-APR-2000; 2000WO-US11459.
XX
PR   30-APR-1999;   99US-0131720.
PR   21-AUG-1999;   99US-0149738.
PR   24-SEP-1999;   99US-0155945.
XX
PA   (BIOS-) BIOSTATUM INC.
XX
PI   Boutaud A;
XX
DR   WPI; 2000-687538/67.
DR   P-PSDB; AAB48461.
XX
PT   Laminin 5-expressing cells, used to accelerate wound healing associated
PT   with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
PT   burns, acute wounds and skin grafts -
XX
PS   Claim 4; Page 110-117; 232pp; English.
XX
CC   The present sequence encodes a laminin 5 chain polypeptide. Recombinant
CC   laminin 5-expressing cells are used to accelerate wound healing,
CC   especially diabetic foot ulcers, venous ulcers, pressure sores, skin
CC   surgery, burns, acute wounds, skin grafts, corneal ulcerations,
CC   gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
CC   used to improve the biocompatibility of medical devices, and to promote
CC   cell adhesion to a surface. They can be used for the ex vivo treatment
CC   of Type I diabetes. Laminin can also be used to regulate angiogensis.
CC   The cell line produces and secretes recombinant heterotrimeric laminin,

CC   whereas prior art cell lines have been created that produce but do not
CC   secreted only one or two chain laminins.
XX
SQ   Sequence 5113 BP; 1449 A; 1245 C; 1331 G; 1088 T; 0 other;

Query Match   6.4%;   Score 28.6; DB 21; Length 5113;
Best Local Similarity 61.3%;   Pred. No. 29;
Matches   46; Conservative   0; Mismatches   0; Indels   29; Gaps   0;

Qy   226 gtgtgctgtttactgtgtagcgcttgatgacttcctgcgtgtactccgcttggagccatt 285
         || ||||| |||||| || || ||| | |||||||| || ||||
Db 1832 GTTTGCTGTAGGCTGTTGAGAGCTGGATTGATTTCTTGCTGTGTAGCCTTTTCCGTGTCATC 1773

Qy   286 tgagtgctgtttacc 300
         | ||||| |
Db 1772 TTGGCCTCGTTTAAC 1758

RESULT 34
AAC83721/c
ID   AAC83721 standard; cDNA; 5264 BP.
XX
AC   AAC83721;
XX
DT   02-MAR-2001 (first entry)
XX
DE   Rat laminin 5 cDNA, SEQ ID NO: 9.
XX
KW   Rat; laminin 5; vulnerary; antiulcer; antiinflammatory; antidiabetic;
KW   cell adhesion promoter; wound healing; ulcers; burn; skin graft;
KW   periodontitis; gingivitis; Type I diabetes; angiogenesis regulation; ss.
XX
OS   Rattus norvegicus.
XX
PN   WO200066731-A2.
XX
PD   09-NOV-2000.
XX
PF   28-APR-2000; 2000WO-US11459.
XX
PR   30-APR-1999;   99US-0131720.
PR   21-AUG-1999;   99US-0149738.
PR   24-SEP-1999;   99US-0155945.
XX
PA   (BIOS-) BIOSTATUM INC.
XX
PI   Boutaud A;
XX
DR   WPI; 2000-687538/67.
DR   P-PSDB; AAB48460.
XX
PT   Laminin 5-expressing cells, used to accelerate wound healing associated
PT   with diabetic foot ulcers, venous ulcers, pressure sores, skin surgery,
PT   burns, acute wounds and skin grafts -
XX
PS   Claim 4; Page 97-104; 232pp; English.
XX
CC   The present sequence encodes a laminin 5 chain polypeptide. Recombinant
CC   laminin 5-expressing cells are used to accelerate wound healing,
CC   especially diabetic foot ulcers, venous ulcers, pressure sores, skin
CC   surgery, burns, acute wounds, skin grafts, corneal ulcerations,
CC   gastro-intestinal ulcers, periodontitis, and gingivitis. They are also
CC   used to improve the biocompatibility of medical devices, and to promote
CC   cell adhesion to a surface. They can be used for the ex vivo treatment
CC   of Type I diabetes. Laminin can also be used to regulate angiogenesis.
CC   The cell line produces and secretes recombinant heterotrimeric laminin,
CC   whereas prior art cell lines have been created that produce but do not
CC   secreted only one or two chain laminins.
XX
SQ   Sequence 5264 BP; 1487 A; 1277 C; 1385 G; 1115 T; 0 other;

Query Match   6.4%;   Score 28.6; DB 21; Length 5264;
Best Local Similarity 61.3%;   Pred. No. 29;
Matches   46; Conservative   0; Mismatches   0; Indels   29; Gaps   0;

Qy   226 gtgtgctgtttactgtgtagcgcttgatgacttcctgcgtgtactccgcttggagccatt 285
         || ||||| |||||| || || ||| | |||||||| || ||||
Db 1983 GTTTGCTGTAGGCTGTTGAGAGCTGGATTGATTTCTTGCTGTGTAGCCTTTTCCGTGTCATC 1924

Qy   286 tgagtgctgtttacc 300
         | ||||| |
Db 1923 TTGGCCTCGTTTAAC 1909

RESULT 35
AAF97897
ID   AAF97897 standard; cDNA; 2288 BP.
XX
AC   AAF97897;
XX
DT   01-JUN-2001 (first entry)
XX
DE   Human secreted protein cDNA, SEQ ID NO: 24.
XX
KW   Human; secreted protein; immunomodulatory; antisclerotic;
KW   dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;
KW   vascular; anti-angiogenic; ophthalmological; neuroprotectant;
KW   nootropic; anticonvulsant; antialzheimers; antiparkinsonian;
KW   antimicrobial; vulnerary; vaccine; gene therapy; cancer;
KW   protein coordinate data; infection; ss.
XX
OS   Homo sapiens.
XX
PN   WO200121658-A1.
XX
PD   29-MAR-2001.
XX
PF   22-SEP-2000; 2000WO-US26013.
XX
PR   24-SEP-1999;   99US-0155709.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;
PI   Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;
PI   Young PE, Wei P, Florence KA;
XX
DR   WPI; 2001-235311/24.
XX
PT   Nucleic acids encoding 32 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 1; Page 729; 890pp; English.
XX
CC   The present sequence encodes one of 32 novel human secreted polypeptides.
CC   The nucleic acid molecules and polypeptides they encode may be used in
CC   the prevention, diagnosis and treatment of diseases such as
CC   immune disorders (e.g. multiple sclerosis, systemic lupus erythematosus
CC   and human immuno-deficiency virus (HIV) infections), hyperproliferative
CC   disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC   (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC   arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC   neovascularisation and diabetic retinopathy), neurological disorders
CC   (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC   infectious diseases and/or for promoting wound healing, regeneration
CC   and/or chemotaxis. The nucleic acid molecules may be used to produce the
CC   secreted polypeptides. They may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acid
CC   sequences in samples. The polypeptides may be used as antigens in the
CC   production of antibodies and in assays to identify modulators of
CC   their expression and activity.
XX
SQ   Sequence 2288 BP; 717 A; 477 C; 490 G; 599 T; 5 other;

```
Query Match    6.4%;    Score 28.4;    DB 22;    Length 2288;
Best Local Similarity  55.2%;    Pred. No. 23;
Matches   53;    Conservative   0;    Mismatches   43;    Indels   0;    Gaps   0;

Qy   226 gtgtgctgttactgtgtagcgcttgatgactccttgctgtactccgcttggaccatt 285
         |||| || |||| |||| |  || ||  |||||| ||||| || || |||||
Db    58 gcgcggggnntccggcgccgaggttcttgactgctgtgcgcggacgccaggtgtagccatg 117

Qy   286 tgagtgctgtttaccgtgcagcactcgaagacttcc 321
         || ||| |||||| ||||| || ||
Db   118 cagcgagccgattccgagcagcagcccctccaagcgtccc 153

RESULT 36
AAF22709/c
ID   AAF22709 standard; cDNA; 815 BP.
XX
AC   AAF22709;
XX
DT   26-MAR-2001  (first entry)
XX
DE   Human gastric cancer associated antigen nucleotide sequence SEQ ID:288.
XX
KW   Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW   cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073801-A2.
XX
PD   07-DEC-2000.
XX
PF   26-MAY-2000; 2000WO-US14749.
XX
PR   28-MAY-1999;   99US-0136526.
PR   10-SEP-1999;   99US-0153454.
XX
PA   (LUDW-) LUDWIG INST CANCER RES.
XX
PI   Obata Y;
XX
DR   WPI; 2001-025274/03.
XX
PT   Nucleic acids encoding breast, gastric and prostate cancer associated
PT   antigen precursors, useful for diagnosing and treating a condition
PT   characterized by expression of an abnormal amount of a protein, e.g.
PT   cancer -
XX
PS   Claim 50; Page 368; 799pp; English.
XX
CC   AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC   represent nucleotide sequences encoding human breast, gastric and
CC   prostate cancer associated antigen precursors (CAAP) respectively,
CC   AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC   represent human breast, gastric and prostate CAAP protein sequence
CC   respectively. CAAPs have cytostatic activity and can be used in the
CC   production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC   acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC   condition characterised by expression of an abnormal amount of a protein,
CC   e.g. cancer.
XX
SQ   Sequence 815 BP; 282 A; 132 C; 123 G; 256 T; 22 other;

Query Match    6.3%;    Score 28;    DB 22;    Length 815;
Best Local Similarity  48.6%;    Pred. No. 19;
Matches   70;    Conservative   0;    Mismatches   74;    Indels   0;    Gaps   0;

Qy   179 gtacaagactaacacacgccaatgggcgtcgaatgttatcggaggccattcgtgctgtttac 238
         ||| |  | ||| ||| || | |||||| | || || | || |||||
Db   642 GTATAAGGCAAGCAAAAGTNTTAACATGGCAGCCATTCTGCNTTTNCNGGGGCCCTGTCC 583


Qy   239 tgtgtagcgcttgatgactccttgctgtactccgcttggaccatttgagtgctgttta 298
         |||| |||| || ||| || |||| || ||||| | |||  ||  || ||||||
Db   582 TGTTTAGTCTTAGTGGGTTAATTTTGTTGTGCGTTAATTTTGCAGAGAAGAAACTTCAGCAAGCTAGA 523

Qy   299 ccgtgcagcactcgaagacttcct 322
         | |||| |||| ||||| ||||
Db   522 ACTGGAAGGACTTTAATTTTCAT 499

RESULT 37
AAA96363
ID   AAA96363 standard; DNA; 50000 BP.
XX
AC   AAA96363;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
XX
KW   Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
KW   ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
KW   insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
KW   Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
KW   thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
KW   Hashimoto's disease; coeliac disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   satellite      5722..5746
FT                  /*tag= a
FT   satellite      6550..6597
FT                  /*tag= b
FT                  /note= "sara41/42 microsatellite repeat"
FT   satellite      19911..19956
FT                  /*tag= c
FT                  /note= "sara43/44 microsatellite repeat"
FT   satellite      23904..23957
FT                  /*tag= d
FT                  /note= "PW210/211 microsatellite repeat"
FT   satellite      27689..27780
FT                  /*tag= e
FT                  /note= "sara45/46 microsatellite repeat"
FT   satellite      30766..30801
FT                  /*tag= f
FT                  /note= "sara17/18 microsatellite repeat"
FT                  /note= "sara19/20 microsatellite repeat"
XX
PN   WO200056856-A2.
XX
PD   28-SEP-2000.
XX
PF   24-MAR-2000; 2000WO-US07938.
XX
PR   25-MAR-1999;   99US-0126215.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Ling V, Wu P, Gray GS;
XX
DR   WPI; 2000-628257/60.
XX
PT   Determining predisposition of humans to develop autoimmune disease
PT   involves detecting polymorphic microsatellite repeat sequence within
PT   human costimulatory receptor gene locus -
XX
PS   Claim 2; Page 67-82; 160pp; English.
XX
CC   Two human bacterial artificial chromosome (BAC) clones that included
CC   and flanked the human CTLA-4 locus were cloned and sequenced. The
CC   sequence data was assembled into a contiguous sequence that is presented
CC   in AAA96363-68. AAA96363-64 comprise BAC clone 22700, and AAA96365-68
```

```
CC   comprise BAC clone 22608. The sequences contain polymorphic
CC   microsatellite repeat (PMR) sequences. The specification describes a
CC   method for determining the predisposition of a human subject to develop
CC   autoimmune disease. The method comprises detecting a PMR sequence in the
CC   CD28, ICOS gene or CTLA4 gene of the human costimulatory receptor gene
CC   locus (hCGRL). PMR sequences vary in length among individuals and can be
CC   amplified to generate products that differ in size. These products can
CC   then be detected by rapid and convenient high resolution processes. The
CC   method is useful for determining the predisposition of insulin-dependent
CC   diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
CC   hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
CC   postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
CC   coeliac disease and leprosy. PMR sequences within hCRGL are useful as
CC   markers in a variety of assays and in the field of forensic medicine,
CC   disease diagnosis and human genome mapping.
XX
SQ   Sequence 50000 BP; 14521 A; 9882 C; 11191 G; 14406 T; 0 other;

Query Match             6.3%;  Score 28;  DB 21;  Length 50000;
Best Local Similarity  50.8%;  Pred. No. 1.4e+02;
Matches   67;  Conservative   0;  Mismatches 65;  Indels  0;  Gaps  0;

Qy    43 atctgaaggaaggacaaagacacctccaagccggacacggctactgtagctggcacggc 102
                   ||| |||| || |||   | | ||||||      |  ||
Db   913 atcacaaggtcaggagatcgagaccatcctggctaacacggtgaaaccccgtctctacta 972

Qy   103 ataatggaaaacagtccctgccctttgagggtcgtctcaaacatgcccctcgagacgtc 162
              | ||| ||||      ||| | || ||| |||| | |||  ||| ||||
Db   973 aaaatacaaaattagccaggcgtggtggcgggcacctgagtcctgctactccagaggct 1032

Qy   163 ggtgaagccgta 174
              | |||  |||
Db  1033 gaggcaggagaa 1044

RESULT 38
AAX57500/c
ID   AAX57500 standard; DNA; 1253 BP.
XX
AC   AAX57500;
XX
DT   27-JUL-1999  (first entry)
XX
DE   Rat Ptx3 cDNA.
XX
KW   Rat; Ptx3; mesencephalic dopaminergic neuron; disorder; neurological;
KW   psychiatric; medication; Parkinson's disease; tardive dyskinesia;
KW   manic depression; schizophrenia; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             64..972
FT                   /tag= a
FT                   /product= "Ptx3"
XX
PN   WO9924572-A1.
XX
PD   20-MAY-1999.
XX
PF   11-NOV-1998;  98WO-NL00652.
XX
PR   11-NOV-1997;  97EP-0203511.
XX
PA   (UVUT-) RIJKSUNIV UTRECHT.
XX
PI   Burbach JPH, Smidt MP;
XX
DR   WPI; 1999-327404/27.
DR   P-PSDB; AAY08461.
XX
PT   Gene with restricted expression in mesencephalic dopaminergic
```

```
PT   neurons
XX
PS   Disclosure; Fig 5; 57pp; English.
XX
CC   This invention describes a novel gene with restricted expression in
CC   mesencephalic dopaminergic neurons, Ptx3, isolated from rat tissue. The
CC   products and methods of the invention can be used to identify a gene
CC   related to neurological or psychiatric disorders. They can be used to
CC   test or develop specific medication for a neurological or psychiatric
CC   disorder, such as Parkinson's disease, tardive dyskinesia, manic
CC   depression, or schizophrenia.
XX
SQ   Sequence 1253 BP; 208 A; 459 C; 390 G; 196 T; 0 other;

Query Match              6.3%;  Score 27.8;  DB 20;  Length 1253;
Best Local Similarity   59.5%;  Pred. No. 27;
Matches   47;  Conservative   0;  Mismatches 32;  Indels  0;  Gaps  0;

Qy    23 gcaggagcgcctagtggtcatctgaaggagaaggacaaaggacaacctccagccggacacg 82
             | || |  |||  |||     ||||| |||| | |||  |  |||| ||||||||||||
Db   809 GCTGCAGCGGCGGCGTCGCGGCCGGCCCGAGGCGTCAAGGCGCAGGACACTCCCCGAGGACACG 750

Qy    83 gctactgtagctggcacgg 101
             |  | || | |||  |||
Db   749 GCGGCTGGAGCCAGCCCGG 731

RESULT 39
AAX59545/c
ID   AAX59545 standard; DNA; 1392 BP.
XX
AC   AAX59545;
XX
DT   21-JUL-1999  (first entry)
XX
DE   Nucleotide sequence of the mouse Pitx3 gene.
XX
KW   Pitx3; homeobox domain protein; lens development; lens disorder;
KW   cataract; detection; ocular disease; ASMD; Peter's anomaly;
KW   anterior segment mesenchymal dysgenesis; ss.
XX
OS   Mus sp.
XX
PN   WO991996-A1.
XX
PD   06-MAY-1999.
XX
PF   26-OCT-1998;   98WO-US22689.
XX
PR   24-OCT-1997;   97US-0957351.
XX
PA   (IOWA ) UNIV IOWA RES FOUND.
XX
PI   Murray JC, Semina EV;
XX
DR   WPI; 1999-312965/26.
DR   P-PSDB; AAY02611.
XX
PT   Pitx3, homeobox protein, and related nucleic acid sequences
XX
PS   Claim 1; Fig 1; 128pp; English.
XX
CC   The present sequence represents the mouse Pitx3 gene. Pitx3 proteins
CC   are homeobox domain proteins, which are involved in the development
CC   of the lens and contribute to diseases and disorders of the lens,
CC   such as cataracts. The Pitx3 nucleic acids (e.g. antisense sequences,
CC   ribozymes and triplex nucleic acids), probes derived from them and
CC   polypeptides, are useful in claimed methods to detect an ocular
CC   disease, especially of the lens, e.g. cataract formation. Specific
CC   conditions that can be detected and treated are Anterior Segment
CC   Mesenchymal Dysgenesis (ASMD) and Peter's anomaly.
XX
```

```
SQ   Sequence 1392 BP; 245 A; 511 C; 426 G; 210 T; 0 other;

Query Match          6.3%;   Score 27.8;   DB 20;   Length 1392;
Best Local Similarity  59.5%;   Pred. No. 28;
Matches   47;  Conservative   0;  Mismatches  32;  Indels  0;  Gaps  0;

Qy    23 gcaggagcgcctagtgggtcatctgaaggaaggacacctccaagccggacacg 82
          || ||||||| |||||||||||| ||| ||||||| ||||||| |||||||||||
Db   879 GCTGCAGCGGCGGCGGCTCGCGGCCGCCGAGGCCGTAAGGCAGGACACTGCCCCGGAGGACACG 820

Qy    83 gctactgtagctggcacgg 101
          || |||||||||||| |||
Db   819 GCGGCTGGAGCGCAGCCCGG 801

RESULT 40
AAX13214/c
ID   AAX13214 standard; DNA; 1663 BP.
XX
AC   AAX13214;
XX
DT   19-MAR-1999   (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:277.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 1295; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   AAX12938 to AAX13919 represent these nucleotide sequences which are
CC   primary nucleotide sequences, also known as contigs. The computer-based
CC   system can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 1663 BP; 543 A; 297 C; 295 G; 525 T; 3 other;

Query Match          6.3%;   Score 27.8;   DB 20;   Length 1663;
```

```
Best Local Similarity  57.5%;   Pred. No. 31;
Matches   50;  Conservative   0;  Mismatches  37;  Indels  0;  Gaps  0;

Qy   213 atcggagcattcgtgtgctgtttactgtgtagcgcttgatgacttcctgctgtactcc 272
          ||||||||:|:||||||| :|:||| :|| ||||||||| ||||| || :||| ||| |
Db   408 ATCGGGAGCGAATAGTGTACCCTTCGCTCTCAATCGTTTGAGAAATGTACTGACGAATAGC 349

Qy   273 gcttggagccattgagtgctgtttac 299
          |||||||| ||||||:|||||||||||
Db   348 CTTTTCATCCATTGGATTGACGTCAAC 322

RESULT 41
AAN91159/c
ID   AAN91159 standard; DNA; 4740 BP.
XX
AC   AAN91159;
XX
DT   07-JUN-1990   (first entry)
XX
DE   DNA sequence downstream of the proteinase promoter fragment from pSK111
DE   in S. cremoris SK211
XX
KW   Lactic acid bacteria; cheese; Streptococcus cremoris SK112; proteinase;
KW   pSK111; ss.
XX
FH   Key              Location/Qualifiers
FT   RBS              49..53
FT                    /*tag= a
FT   signal_peptide   61..159
FT                    /*tag= b
FT   mature_protein   160..4740
FT                    /*tag= c
FT                    product= proteinase
XX
PN   NL8701378-A.
XX
PD   02-JAN-1989.
XX
PF   12-JUN-1987;   87NL-0001378.
XX
PR   12-JUN-1987;   87NL-0001378.
XX
PA   (NEZU-) NEDERL INS ZUIVELON.
XX
PI   Simons AFM,  De Vos WM;
XX
DR   WPI; 1989-030097/04.
DR   P-PSDB; AAP94145.
XX
PT   DNA fragment having region specific for lactic acid bacteria - is
PT   contained in plasmid in microorganism used in prodn. of protein and
PT   food prodn. eg cheese.
XX
PS   Claim 6; fig 7a; 43pp; Dutch.
XX
CC   DNA encoding prochymosin can be cloned into a plasmid ( esp. from
CC   S. cremoris SK112) and used to produce large amts of the protein by
CC   recombinant DNA techniques. The plasmid pref. contains this sequence
CC   from the downstream region of the proteinase regulatory region(used to
CC   control expression of the protein). Prodn of prochymosin in this way
CC   could overcome the shortage of prochymosin due to a shortage of calf
CC   stomachs and increasing cheese prodn. Prochymosin is also used in prodn.
CC   of yoghurt, butter and buttermilk.
CC   See also AAN91157-N91160.
XX
SQ   Sequence 4740 BP; 1424 A; 1095 C; 1099 G; 1122 T; 0 other;

Query Match          6.3%;   Score 27.8;   DB 10;   Length 4740;
Best Local Similarity  53.2%;   Pred. No. 51;
Matches   59;  Conservative   0;  Mismatches  52;  Indels  0;  Gaps  0;
```

```
Qy    75 cggacacggctactgtagctgcacggcataatggaaaacagtccctgcccttttgaggg 134
         ||| || ||||| | ||| ||| || |  ||| |      ||| ||||||||| |||| ||
Db  3817 CGGTGTGGCCAACATTAACTGTCTTGGTGTCAGCACTAACCGTCCCTGTAACCTTGATGG 3758

Qy   135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaag 185
         ||| ||| ||  |||| ||| |  ||||  ||||| |   || ||| |||
Db  3757 TCGCTTCAGAGGTCTGATCCGAGCCATTGTTGAATTTCAAATCAGTAAAGG 3707
```

RESULT 42
AAQ10416/c
ID   AAQ10416 standard; DNA; 6666 BP.
XX
AC   AAQ10416;
XX
DT   15-APR-1991 (first entry)
XX
DE   Mutant protease gene (delta137-139).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria; ss.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key   Location/Qualifiers
FT   CDS   376..6253
FT   /*tag= a
FT   /product= mutant protease
FT   mat_peptide   938..6253
FT   /*tag= b
FT   sig_peptide   376..937
FT   /*tag= c
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   P-PSDB; AAR10562.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure: Fig 1(1-7)+5(b); 29 pp; English.
XX
CC   The wild-type L.lactis SK11 protease gene sequence was determined
CC   by the applicant (EP-307011).
CC   The mutant protease having mew cleavage specificities is
CC   obtained by deleting three amino acids (nine bps).
CC   This mutant may then be used to prepare hybrid protease genes,
CC   the fusion being between a type I and a type III protease gene of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s). The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence 6666 BP; 2005 A; 1499 C; 1553 G; 1609 T; 0 other;

Query Match   6.3%;   Score 27.8; DB 12; Length 6666;
Best Local Similarity 53.2%;   Pred. No. 60;

Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy    75 cggacacggctactgtagctgcacggcataatggaaaacagtccctgcccttttgaggg 134
         ||| || ||||| | ||| ||| || |  ||| |      ||| ||||||||| |||| ||
Db  4123 CGGTGTGGCCAACATTAACTGTCTTGGTGTCAGCACTAACCGTCCCTGTAACCTTGATGG 4064

Qy   135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaag 185
         ||| ||| ||  |||| ||| |  ||||  ||||| |   || ||| |||
Db  4063 TCGCTTCAGAGGTCTGATCCGAGCCATTGTTGAATTTCAAATCAGTAAAGG 4013
```

RESULT 43
AAQ10414/c
ID   AAQ10414 standard; DNA; 6675 BP.
XX
AC   AAQ10414;
XX
DT   15-APR-1991 (first entry)
XX
DE   Mutant protease gene (K138N).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria; ss.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key   Location/Qualifiers
FT   CDS   376..6262
FT   /*tag= a
FT   /product= mutant protease
FT   mat_peptide   938..6262
FT   /*tag= b
FT   sig_peptide   376..937
FT   /*tag= c
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990;   90EP-0202113.
XX
PR   04-AUG-1989;   89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   P-PSDB; AAR10560.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure: Fig 1(1-7)+5(a); 29 pp; English.
XX
CC   The wild-type L.lactis SK11 protease gene sequence was determined
CC   by the applicant (EP-307011).
CC   The mutant protease K138N having mew cleavage specificities is
CC   obtained by carrying out single amino acid substitutions.
CC   This mutant may then be used to prepare hybrid protease genes,
CC   the fusion being between a type I and a type III protease gene of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s). The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence 6675 BP; 2008 A; 1502 C; 1555 G; 1610 T; 0 other;

Query Match   6.3%;   Score 27.8; DB 12; Length 6675;
Best Local Similarity   53.2%;   Pred. No. 60;
Matches 59; Conservative   0; Mismatches 52; Indels   0; Gaps   0;

Qy   75 cggacacggctactgtagctggcacggcataatggaaaacagtcctgccttttgaggg 134
           |||.||.||||.||.|||.|| ||||||||||||||.||||| |||||||||||||| ||
Db   4132 CGGTGTGGCCAACATTAACTGTCTTGGTGTGTCAGCACTAACCGTCCCTGTAACCTTGATGG 4073

Qy   135 tcgtctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaag 185
           || ||||.|| |||.|||||||||||||| || |||||||| ||||||||||
Db   4072 TCGCTTCAGAGGTCTGATCCGAGCCATTGTTGAATTTCAAATCAGTAAAGG 4022

RESULT 44
AAQ10415/c
ID   AAQ10415 standard; DNA; 6675 BP.
XX
AC   AAQ10415;
XX
DT   15-APR-1991 (first entry)
XX
DE   Mutant protease gene (N166D).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria; ss.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key   Location/Qualifiers
FT   CDS   376..6262
FT   /*tag= a
FT   /product= mutant protease
FT   mat_peptide   938..6262
FT   /*tag= b
FT   sig_peptide   376..937
FT   /*tag= c
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990; 90EP-0202113.
XX
PR   04-AUG-1989; 89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   P-PSDB; AAR10561.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC   The wild-type L.lactis SK11 protease gene sequence was determined
CC   by the applicant (EP-307011).
CC   The mutant protease N166D having mew cleavage specificities is
CC   obtained by carrying out single amino acid substitution.
CC   This mutant may then be used to prepare hybrid protease genes,
CC   the fusion being between a type I and a type III protease gene of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s). The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.
XX
SQ   Sequence 6675 BP; 2007 A; 1502 C; 1556 G; 1610 T; 0 other;

Query Match   6.3%;   Score 27.8; DB 12; Length 6675;
Best Local Similarity   53.2%;   Pred. No. 60;
Matches 59; Conservative   0; Mismatches 52; Indels   0; Gaps   0;

Qy   75 cggacacggctactgtagctggcacggcataatggaaaacagtcccgccttttgaggg 134
           |||.||.||||.||.|||.|| ||||||||||||||.||||| |||||||||||||| ||
Db   4132 CGGTGTGGCCAACATTAACTGTCTTGGTGTGTCAGCACTAACCGTCCCTGTAACCTTGATGG 4073

Qy   135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaag 185
           || ||||.|| |||.||||||||||||| || |||||||| ||||||||||
Db   4072 TCGCTTCAGAGGTCTGATCCGAGCCATTGTTGAATTTCAAATCAGTAAAGG 4022

RESULT 45
AAQ10411/c
ID   AAQ10411 standard; DNA; 6675 BP.
XX
AC   AAQ10411;
XX
DT   15-APR-1991 (first entry)
XX
DE   Mutant protease gene (A137G/K138D).
XX
KW   Mutant protease gene; fermentation; foodstuff; flavouring;
KW   lactic acid bacteria; ss.
XX
OS   Lactococcus lactis SK11.
XX
FH   Key   Location/Qualifiers
FT   CDS   376..6262
FT   /*tag= a
FT   /product= mutant protease
FT   mat_peptide   938..6262
FT   /*tag= b
FT   sig_peptide   376..937
FT   /*tag= c
XX
PN   EP411715-A.
XX
PD   06-FEB-1991.
XX
PF   02-AUG-1990; 90EP-0202113.
XX
PR   04-AUG-1989; 89NL-0002010.
XX
PA   (NEZU-) NED INST ZUIVELONDE.
XX
PI   Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
XX
DR   WPI; 1991-038622/06.
DR   P-PSDB; AAR10557.
XX
PT   Mutant protease gene(s) and protease(s) - derived from type I and
PT   III protease genes from lactococcal strains, used in fermentation
PT   foodstuffs and flavourings
XX
PS   Disclosure; Fig 1(1-7)+5(b); 29 pp; English.
XX
CC   The wild-type L.lactis SK11 protease gene sequence was determined
CC   by the applicant (EP-307011).
CC   The mutant protease A137G/K138D having mew cleavage specificities is
CC   obtained by replacing two amino acids (nine bps).
CC   This mutant may then be used to prepare hybrid protease genes,
CC   the fusion being between a type I and a type III protease gene of
CC   L.lactis Wg2 and SK11.
CC   The product has modified properties, e.g. thermostability,
CC   alkaline/acid pH stability, oxidative stability, autoproteolysis etc.,
CC   compared to the parent protease(s). The proteases can be used for
CC   preparing products (butter cheese, human and animal foodstuffs)
CC   prepared with the aid of lactic acid bacteria.
CC   See also AAQ10411-17 and AAQ10870-71.

XX
SQ   Sequence 6675 BP; 2007 A; 1502 C; 1557 G; 1609 T; 0 other;

Query Match   6.3%;   Score 27.8;   DB 12;   Length 6675;
Best Local Similarity 53.2%;   Pred. No. 60;
Matches   59;   Conservative   0;   Mismatches   52;   Indels   0;   Gaps   0;

Qy   75 cggacacggctactgtgagctggcacgcataatggaaaacagtccctgccctttgaggg 134
          ||| ||| | | || |||| | | | | | | || |||| |||| | ||||||||
Db   4132 CGGTGTGGCCAACATTAACTGTCTTGGTGTCAGCACTAACCGTCCCTGTAACCTTGATGG 4073

Qy   135 tcgtctcaaacatgccctcgagacgtcggtgaagccgtagtttgtacaag 185
          ||| ||| | | | || | | | ||||| | ||| | ||| | ||| || |
Db   4072 TCGCTTCAGAGGTCTGATCCGAGCCATTGTTGAATTTCAAATCAGTAAAGG 4022

Search completed: September 7, 2001, 16:41:35
Job.time: 1509 sec

THIS PAGE BLANK (USPTO)

OM nucleic - nucleic search, using sw model

Copyright GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: September 7, 2001, 14:58:56 ; Search time 1472.14 Seconds
                 (without alignments)
                 4665.100 Million cell updates/sec

Title: US-09-540-234-1
Perfect score: 444
Sequence: 1 gggtccacggcccttagt........tgtactcgtctgttcgag 444

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
            1:   gb_ba1:*
            2:   gb_ba2:*
            3:   gb_ba3:*
            4:   gb_in1:*
            5:   gb_in2:*
            6:   gb_in3:*
            7:   gb_om:*
            8:   gb_ov:*
            9:   gb_pat1:*
           10:   gb_pat2:*
           11:   gb_ph:*
           12:   gb_pl1:*
           13:   gb_pl2:*
           14:   gb_pl3:*
           15:   gb_pl4:*
           16:   em_ba1:*
           17:   em_ba2:*
           18:   em_fun:*
           19:   em_htgo_hum:*
           20:   em_htgo_inv:*
           21:   em_htgo_rod:*
           22:   em_htg_hum1:*
           23:   em_htg_hum2:*
           24:   em_htg_hum3:*
           25:   em_htg_hum4:*
           26:   em_htg_hum5:*
           27:   em_htg_hum6:*
           28:   em_htg_hum7:*
           29:   em_htg_hum8:*
           30:   em_htg_inv1:*
           31:   em_htg_inv2:*
           32:   em_htg_other:*
           33:   em_htg_rod:*
           34:   em_hum1:*
           35:   em_hum2:*
           36:   em_hum3:*
           37:   em_hum4:*
           38:   em_hum5:*
           39:   em_hum6:*
           40:   em_hum7:*
           41:   em_in:*
           42:   em_om:*
           43:   em_or:*
           44:   em_ov:*
           45:   em_pat:*
           46:   em_ph:*
           47:   em_pl:*
           48:   em_ro:*
           49:   em_sts:*
           50:   em_sy:*
           51:   em_un:*
           52:   em_vi:*
           53:   gb_sts1:*
           54:   gb_sts2:*
           55:   gb_sts3:*
           56:   gb_sy:*
           57:   gb_un:*
           58:   gb_vil:*
           59:   gb_vi2:*
           60:   gb_htg1:*
           61:   gb_htg2:*
           62:   gb_htg3:*
           63:   gb_htg4:*
           64:   gb_htg5:*
           65:   gb_htg6:*
           66:   gb_htg7:*
           67:   gb_htg8:*
           68:   gb_htg9:*
           69:   gb_htg10:*
           70:   gb_htg11:*
           71:   gb_htg12:*
           72:   gb_htg13:*
           73:   gb_htg14:*
           74:   gb_htg15:*
           75:   gb_htg16:*
           76:   gb_htg17:*
           77:   gb_htg18:*
           78:   gb_htg19:*
           79:   gb_htg20:*
           80:   gb_htg21:*
           81:   gb_htg22:*
           82:   gb_htg23:*
           83:   gb_htg24:*
           84:   gb_htg25:*
           85:   gb_pr1:*
           86:   gb_pr2:*
           87:   gb_pr3:*
           88:   gb_pr4:*
           89:   gb_pr5:*
           90:   gb_pr6:*
           91:   gb_pr7:*
           92:   gb_pr8:*
           93:   gb_pr9:*
           94:   gb_rol:*
           95:   gb_ro2:*
           96:   gb_in4:*
           97:   gb_pr10:*
           98:   em_ba3:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID		Description
1	41.6	9.4	173843	77	AC084405	AC084405	Oryza sat
2	37.2	8.4	166723	75	AC074278	AC074278	Homo sapi
3	36.8	8.3	164289	76	AC079395	AC079395	Homo sapi
4	36.8	8.3	193308	81	AL391703	AL391703	Homo sapi
5	35.6	8.0	166847	84	CNS06C7S	AL390801	Homo sapi
6	35.6	8.0	167068	84	CNS06C88	AL391516	Homo sapi
7	35.6	8.0	176506	84	CNS06C7O	AL390335	Homo sapi
8	35.2	7.9	1042	2	BCHMLYSN	X84058	B.cereus ge

```
RESULT 1
AC084405
LOCUS       AC084405    173843 bp    DNA                    HTG       23-MAR-2001
DEFINITION  Oryza sativa chromosome unknown clone OSJNBa0044H10, *** SEQUENCING
            IN PROGRESS ***, 2 unordered pieces.
ACCESSION   AC084405
VERSION     AC084405.3  GI:13435260
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 173843)
  AUTHORS   Buell,R., Hsiao,J., Zismann,V., Moffat,K.M., Hill,J.,
            Gansberger,K., Burgess,S., Jarrahi,B., Shvartsbeyn,M., Brenner,M.,
            Ciecko,A., Pai,G., Vanaken,S., Hansen,C., Utterback,T.,
            Feldblyum,T., Khalak,H.G., Yuan,Q., Quackenbush,J., White,O.,
            Salzberg,S. and Fraser,C.
  TITLE     Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0044H10 BAC genomic
            sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173843)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Mar 23, 2001 this sequence version replaced gi:11968424.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
                                              ALIGNMENTS
```

```
c   9   34.8   7.8     7218       10   I66494       10   I66494 Sequence 14
c  10   34.8   7.8   180944       75   AC073824      75   AC073824 Mus muscu
c  11   34.6   7.8   113836       89   AL133174      89   AL133174 Human DNA
c  12   34.2   7.7      397        7   AF123274       7   AF123274 Physeter
c  13   34.2   7.7   193583       73   AC068473      73   AC068473 Homo sapi
   14   34     7.7   152393       66   AC021765      66   AC021765 Homo sapi
c  15   34     7.7   158113       62   AC011159      62   AC011159 Homo sapi
c  16   34     7.7   174059       81   AL512489      81   AL512489 Homo sapi
   17   33.8   7.6   172056       70   AC026746      70   AC026746 Homo sapi
   18   33.8   7.6   179556       79   AL162256      79   AL162256 Homo sapi
   19   33.8   7.6   189470       64   AC016645      64   AC016645 Homo sapi
c  20   33.8   7.6   189942       74   AC073540      74   AC073540 Homo sapi
c  21   33.8   7.6   196716       70   AC026750      70   AC026750 Homo sapi
   22   33.8   7.6   213351       70   AC026809      70   AC026809 Homo sapi
   23   33.6   7.6      273       97   HUMF13A14     97   HUMF13A14
   24   33.6   7.6   150349       67   AC023009      67   AC023009 Human coagu
   25   33.6   7.6   163960       90   AL157775      90   AL157775 Human DNA
c  26   33.6   7.6   166402       69   AC025972      69   AC025972 Homo sapi
   27   33.6   7.6   188642       73   AC068860      73   AC068860 Homo sapi
   28   33.6   7.6   281423       78   AL139335      78   AL139335 Homo sapi
c  29   33.4   7.5     1894        9   AX048148       9   AX048148 Sequence
c  30   33.4   7.5     1894       88   AF071538      88   AF071538 Homo sapi
c  31   33.4   7.5     1898       85   AB031549      85   AB031549 Homo sapi
c  32   33.4   7.5     3317        9   AX048159       9   AX048159 Sequence
   33   33.4   7.5    65534       73   AC068504      73   AC068504 Homo sapi
c  34   33.4   7.5   200729       67   AC022781      67   AC022781 Mus muscu
c  35   33.2   7.5   128219       80   AL356364      80   AL356364 Homo sapi
c  36   33.2   7.5   180381       81   AL513283      81   AL513283 Homo sapi
   37   32.8   7.4   209599       68   AC023785      68   AC023785 Homo sapi
c  38   32.8   7.4   217060       82   AL589931      82   AL589931 Homo sapi
c  39   32.6   7.3   157890       68   AC023681      68   AC023681 Drosophil
c  40   32.6   7.3   177407       78   AL137002      78   AL137002 Homo sapi
c  41   32.6   7.3   313750       79   AL162454      79   AL162454 Homo sapi
c  42   32.4   7.3     1983        8   CLU20353       8   CLU20353 Columba liv
   43   32.4   7.3     2263       14   FRX297967     14   AJ297967 Fragaria
c  44   32.4   7.3   186053       65   AC018863      65   AC018863 Homo sapi
c  45   32.4   7.3   190650       63   AC013799      63   AC013799 Homo sapi
```

```
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1   145511: contig of 145511 bp in length
*                gap of unknown length
*   145512 173843: Contig of 28332 bp in length.
FEATURES             Location/Qualifiers
     source          1..173843
                     /organism="Oryza sativa"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:4530"
                     /chromosome="unknown"
                     /clone="OS38-OSJNBa0044H10"
                     /clone="OSJNBa0044H10"
BASE COUNT    49108 a    38285 c    38525 g    47925 t
ORIGIN
```

```
Query Match                9.4%;      Score 41.6;  DB 77;    Length 173843;
Best Local Similarity     67.3%;      Pred. No. 0.066;
Matches  74;   Conservative   0;   Mismatches  34;   Indels   2;   Gaps   1;

Qy    138  tctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaagactaacaacgcc  197
           |||||||| |||||||||||||  |||| ||||| ||||| ||||||||||| ||||||||
Db  97745  TCTCGGACATGGCCCTTACTTTCGGGTGAAGCCGTAGTTGTACAACACTAACGACATC  97804

Qy    198  aatggcgtc--gatgttatcggagccattcgtgtgctgtttactgtgtag  245
           |||| | ||  ||||| |||||||| |||||||  |||| |||||||||||
Db  97805  AGTGGCATCATCATGTTTTTCCTAGCCATGTTGAGTGCTATTACTGTGTTG  97854
```

```
RESULT 2
AC074278
LOCUS       AC074278    166723 bp    DNA                    HTG       05-SEP-2000
DEFINITION  Homo sapiens chromosome 3q clone RP11-253B24, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC074278
VERSION     AC074278.10  GI:9966094
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166723)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,N., Nguyen,R., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,C., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166723)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

COMMENT   Baylor Plaza, Houston, TX 77030, USA
            On Sep 4, 2000 this sequence version replaced gi:9964664.
            ------ Genome Center ------
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ------ Project Information ------
            Center project name: HBRF
            Center clone name: RP11-253B24
            ------ Summary Statistics ------
            Sequencing vector: M13; L08821
            Chemistry: Dye-primer Bodipy: 40% of reads
            Chemistry: Dye-terminator Big Dye: 59% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 157482 bases at least Q40
            Consensus quality: 161244 bases at least Q30
            Consensus quality: 163250 bases at least Q20
            Estimated insert size: 163632; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 6.1x in Q20 bases; sum-of-contigs estimation
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 12 contigs. The true order of the pieces
          *   is not known and their order in this sequence record is
          *   arbitrary. Gaps between the contigs are represented as
          *   runs of N, but the exact sizes of the gaps are unknown.
          *   This record will be updated with the finished sequence
          *   as soon as it is available and the accession number will
          *   be preserved.
          *   1   42137: contig of 42137 bp in length
          *   42138   42237: gap of unknown length
          *   42238   70998: contig of 28761 bp in length
          *   70999   71098: gap of unknown length
          *   71099   89234: contig of 18136 bp in length
          *   89235   89334: gap of unknown length
          *   89335   100935: contig of 11601 bp in length
          *   100936   101035: gap of unknown length
          *   101036   118663: contig of 17628 bp in length
          *   118664   118763: gap of unknown length
          *   118764   131094: contig of 12331 bp in length
          *   131095   131194: gap of unknown length
          *   131195   140678: contig of 9484 bp in length
          *   140679   140778: gap of unknown length
          *   140779   152452: contig of 11674 bp in length
          *   152453   152552: gap of unknown length
          *   152553   158042: contig of 5490 bp in length
          *   158043   158142: gap of unknown length
          *   158143   163036: contig of 4894 bp in length
          *   163037   163136: gap of unknown length
          *   163137   165583: contig of 2447 bp in length
          *   165584   165683: gap of unknown length
          *   165684   166723: contig of 1040 bp in length.
FEATURES   Location/Qualifiers
     source   1..166723
                     /organism"Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3q"
                     /clone="RP11-253B24"
BASE COUNT   46299 a   36483 c   37271 g   45548 t   1122 others
ORIGIN

Query Match   8.4%; Score 37.2; DB 75; Length 166723;
Best Local Similarity 48.6%; Pred. No. 1.5;
Matches 102; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Qy   216 ggagccattcggtgtgcttgttactgtgtagccgttgatgacttcctttgctgtactccgct 275
           ||||| || |||| || ||| | ||||||||| ||||||| || |||| || |||| ||||
Db 161842 GCAGCCTTTGCCTGTATGGAATTCTTCTTGGTGCATCAGCAGTTCTCGCTGCTTCTCCTAT 161901

Qy   276 tgggagccattggtgctgttaccgtgcagcactcgaagacttcctgctgtactcgct 335
           |||| ||| || ||||| ||||| || || |||| |||| |||| ||| |||||| |||
Db 161902 TTTATTCTTACCATCCTATGGACTTTGCACACTAAAACGTTTGCTTGGTTTAGTTGGT 161961

Qy   336 ggcagcatagtctgttgctgccggtctgtgggacaccgtgtactacagtacttcgttt 395
           |||| || | || ||| || |||||| || || |||| ||||| || ||| |||| |||
Db 161962 AGAGTTTCTTGTCTTTGGCAAATGTTCCAGACTCATCTGAGATCCCTAAGTAAAAGTTT 162021

Qy   396 cccttgtagtggtactacttcttagccat 425
           |||| | ||||| |||| |||| ||| ||
Db 162022 GTTTAAAGAGCAATTTAATTCTTGACCCT 162051

RESULT   3
AC079395/c
LOCUS   AC079395   164289 bp   DNA   HTG   12-OCT-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-219H23, WORKING DRAFT
            SEQUENCE, 34 unordered pieces.
ACCESSION   AC079395
VERSION   AC079395.3 GI:10799480
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 164289)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 164289)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-AUG-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Oct 12, 2000 this sequence version replaced gi:9964991.
            ------ Genome Center ------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ------ Project Information ------
            Center project name: H_NH0219H23
            ------ Summary Statistics ------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET: 100% of reads
            Chemistry: Dye-terminator Big Dye: 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 144473 bases at least Q40
            Consensus quality: 151522 bases at least Q30
            Consensus quality: 154634 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 160989; sum-of-contigs
            Quality coverage: 3.04 in Q20 bases; agarose-fp
            Quality coverage: 3.33 in Q20 bases; sum-of-contigs
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 34 contigs. The true order of the pieces
          *   is not known and their order in this sequence record is
          *   arbitrary. Gaps between the contigs are represented as
          *   runs of N, but the exact sizes of the gaps are unknown.
          *   This record will be updated with the finished sequence
          *   as soon as it is available and the accession number will
          *   be preserved.
          *   1   1493: contig of 1493 bp in length
          *   1494   1593: gap of unknown length
          *   1594   2880: contig of 1287 bp in length
          *   2881   2980: gap of unknown length
          *   2981   4341: contig of 1361 bp in length
          *   4342   4441: gap of unknown length
          *   4442   5719: contig of 1278 bp in length
          *   5720   5819: gap of unknown length

```
*      5820      7200: contig of 1381 bp in length
*      7201      7300: gap of unknown length
*      7301      8864: contig of 1564 bp in length
*      8865      8964: gap of unknown length
*      8965     10832: contig of 1868 bp in length
*     10833     10932: gap of unknown length
*     10933     12176: contig of 1244 bp in length
*     12177     12276: gap of unknown length
*     12277     13636: contig of 1360 bp in length
*     13637     13736: gap of unknown length
*     13737     15321: contig of 1585 bp in length
*     15322     15421: gap of unknown length
*     15422     17156: contig of 1735 bp in length
*     17157     17256: gap of unknown length
*     17257     19049: contig of 1793 bp in length
*     19050     19149: gap of unknown length
*     19150     22210: contig of 3061 bp in length
*     22211     22310: gap of unknown length
*     22311     23900: contig of 1590 bp in length
*     23901     24000: gap of unknown length
*     24001     26618: contig of 2618 bp in length
*     26619     26718: gap of unknown length
*     26719     29184: contig of 2466 bp in length
*     29185     29284: gap of unknown length
*     29285     32405: contig of 3121 bp in length
*     32406     32505: gap of unknown length
*     32506     35754: contig of 3249 bp in length
*     35755     35854: gap of unknown length
*     35855     38675: contig of 2821 bp in length
*     38676     38775: gap of unknown length
*     38776     42794: contig of 4019 bp in length
*     42795     42894: gap of unknown length
*     42895     48461: contig of 5567 bp in length
*     48462     48561: gap of unknown length
*     48562     52170: contig of 3609 bp in length
*     52171     52270: gap of unknown length
*     52271     58241: contig of 5971 bp in length
*     58242     58341: gap of unknown length
*     58342     63672: contig of 5331 bp in length
*     63673     63772: gap of unknown length
*     63773     71246: contig of 7474 bp in length
*     71247     71346: gap of unknown length
*     71347     79238: contig of 7892 bp in length
*     79239     79338: gap of unknown length
*     79339     84766: contig of 5428 bp in length
*     84767     84866: gap of unknown length
*     84867     96100: contig of 11234 bp in length
*     96101     96200: gap of unknown length
*     96201    103109: contig of 6909 bp in length
*    103110    103209: gap of unknown length
*    103210    111868: contig of 8659 bp in length
*    111869    111968: gap of unknown length
*    111969    124202: contig of 12234 bp in length
*    124203    124302: gap of unknown length
*    124303    137678: contig of 13376 bp in length
*    137679    137778: gap of unknown length
*    137779    149722: contig of 11944 bp in length
*    149723    149822: gap of unknown length
*    149823    164289: contig of 14467 bp in length.

FEATURES             Location/Qualifiers
     source          1..164289
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-219H23"
     misc_feature    1..1493
                     /note="assembly_name:Contig11"
     misc_feature    1594..2880
                     /note="assembly_name:Contig12"
     misc_feature    2981..4341
                     /note="assembly_name:Contig13"
     misc_feature    4442..5719
                     /note="assembly_name:Contig14"

     misc_feature    5820..7200
                     /note="assembly_name:Contig15"
     misc_feature    7301..8864
                     /note="assembly_name:Contig16"
     misc_feature    8965..10832
                     /note="assembly_name:Contig17"
     misc_feature    10933..12176
                     /note="assembly_name:Contig18"
     misc_feature    12277..13636
                     /note="assembly_name:Contig19"
     misc_feature    13737..15321
                     /note="assembly_name:Contig20"
     misc_feature    15422..17156
                     /note="assembly_name:Contig21"
     misc_feature    17257..19049
                     /note="assembly_name:Contig22"
     misc_feature    19150..22210
                     /note="assembly_name:Contig23"
     misc_feature    22311..23900
                     /note="assembly_name:Contig24"
     misc_feature    24001..26618
                     /note="assembly_name:Contig25"
     misc_feature    26719..29184
                     /note="assembly_name:Contig26"
     misc_feature    29285..32405
                     /note="assembly_name:Contig27"
     misc_feature    32506..35754
                     /note="assembly_name:Contig28"
     misc_feature    35855..38675
                     /note="assembly_name:Contig29
                     clone_end:SP6
                     vector_side:right"
     misc_feature    38776..42794
                     /note="assembly_name:Contig30"
     misc_feature    42895..48461
                     /note="assembly_name:Contig31"
     misc_feature    48562..52170
                     /note="assembly_name:Contig32"
     misc_feature    52271..58241
                     /note="assembly_name:Contig33"
     misc_feature    58342..63672
                     /note="assembly_name:Contig34"
     misc_feature    63773..71246
                     /note="assembly_name:Contig35"
     misc_feature    71347..79238
                     /note="assembly_name:Contig36"
     misc_feature    79339..84766
                     /note="assembly_name:Contig37"
     misc_feature    84867..96100
                     /note="assembly_name:Contig38"
     misc_feature    96201..103109
                     /note="assembly_name:Contig39"
     misc_feature    103210..111868
                     /note="assembly_name:Contig40"
     misc_feature    111969..124202
                     /note="assembly_name:Contig41"
     misc_feature    124303..137678
                     /note="assembly_name:Contig42"
     misc_feature    137779..149722
                     /note="assembly_name:Contig43"
     misc_feature    149823..164289
                     /note="assembly_name:Contig44"
BASE COUNT     44379 a   36830 c   36917 g   42838 t    3325 others
ORIGIN
```

```
Query Match                  8.3%;    Score 36.8;   DB 76;   Length 164289;
Best Local Similarity       48.6%;    Pred. No. 2;
Matches  101;  Conservative   0;  Mismatches  107;  Indels   0;  Gaps   0;

Qy    216 ggagccattcgtgtcgtgtttactgtgtagccgttgatgactccgtctgctgactccgct  275
Db  27886 GCAGCCTTGCTTGCTGTATGGAATTCTTCTTGGTGCATCAGCAGTTTCTCGCTGCTTCTCCTAT 27827
```

```
Qy    276  tggagccattgagtgctgttaccgtgcagcactcgaagacttcctgctgtactcgct  335
                || ||| ||||||||||||| |||||  ||| ||  |||| |||||||||| ||
Db  27826  TTTATTCTTACCATCCCTATGGACTTTGCAACACTAAAACGTTGCTTGGTTTAGTTGGT  27767

Qy    336  ggcagcatagtcgtttgctgcctggtctgggacacctgtgtactacagtacttcgttt  395
                || ||||| ||||||||||  ||| |||| ||  |||||  ||   ||||
Db  27766  AGAGTTTCTTGCTTTGGCAAATGTTCCAGACTCATCTGAGATCCCTAAGTAAAAGTTT  27707

Qy    396  ccctttgtagtggtactacttcttagcc  423
                || |||  || ||||||| ||
Db  27706  GTTTAAAGAGCAATTTTAATTCTTGCCC  27679


RESULT       4
AL391703
LOCUS        AL391703    193308 bp    DNA                   HTG       08-APR-2001
DEFINITION   Homo sapiens chromosome 1 clone RP11-422P8, *** SEQUENCING IN
             PROGRESS ***, 3 unordered pieces.
ACCESSION    AL391703
VERSION      AL391703.5  GI:13568198
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 193308)
  AUTHORS    Plumb,B.
  TITLE      Direct Submission
  JOURNAL    Submitted (08-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
             requests: clonerequest@sanger.ac.uk
             On Apr 9, 2001 this sequence version replaced gi:10178826.
COMMENT      ------------- Genome Center
             Center: Sanger Centre
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: humquery@sanger.ac.uk
             ------------- Project Information
             Center project name: bA422P8
             ------------- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Sequencing vector: plasmid; L08752; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Consensus quality: 192613 bases at least Q40
             Consensus quality: 192863 bases at least Q30
             Consensus quality: 192997 bases at least Q20
             Insert size: 193108; sum-of-contigs
             Insert size: 194753; 8.7% error; agarose-fp
             Quality coverage: 7.89x in Q20 bases; sum-of-contigs Quality
             coverage: 7.87x in Q20 bases; agarose-fp
             -------------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 3 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     40476: contig of 40476 bp in length
             *   40477  40576: gap of    100 bp
             *   40577 181999: contig of 141423 bp in length
             *  182000 182099: gap of    100 bp
             *  182100 193308: contig of 11209 bp in length.
FEATURES             Location/Qualifiers
     source          1..193308
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-422P8"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..40476
```

```
                     /note="assembly_fragment:01657
                     fragment_chain:1"
     misc_feature    40577..181999
                     /note="assembly_fragment:03665
                     fragment_chain:1"
     misc_feature    182100..193308
                     /note="assembly_fragment:03205
                     fragment_chain:1"
BASE COUNT   56652 a   38556 c   38995 g   58904 t     201 others
ORIGIN

Query Match                       8.3%;  Score 36.8;  DB 81;  Length 193308;
Best Local Similarity   48.1%;  Pred. No. 2;
Matches 104;  Conservative   0;  Mismatches 112;  Indels   0;  Gaps   0;

Qy    200  tggcgtgcatgttatcggagccattcgtgctgcttgttactgtgtagcgcttgatgacttc  259
                |||| |||||||| ||||||| ||||||||| |||||||||||| ||||||||
Db 121135  TGTTGTCTCATGATTTTTCAGCAGTGTTTTCTAGTTTTTCTTGTAGAGGTATTTCACTTC  121194

Qy    260  cttgctgtactccgcttggagccattggagtgctgtttaccgtgcagcactcgaagactt  319
                |||||| || ||||||  |  |||||||   ||||||||  ||||||||||| |||||
Db 121195  CTTGGTTAGGTATATTCCTAAGCATTTTTTATTTTATTTTGCAGCTGTTGTAAAAGG  121254

Qy    320  cctgtctgtactgcgcgagcatagtctgtttgcctgggacacctgtgta  379
                |||  ||| | ||| || ||| || |||||||| |||||||| | |||
Db 121255  GGTTGAGTTCTTGACTTGATTCTCAGCTGTTGGCTGTGTGGTGTATAGGAGAGCTACTGA  121314

Qy    380  ctacagtacttcgtttccccttgtagtggtactact  415
                |  ||| |||| |||||| |||||||  |||
Db 121315  TTTCGCGTACATTAATTTTGTCTCGAAACGTTACT  121350


RESULT       5
CNS06C7S
LOCUS        CNS06C7S    166847 bp    DNA                   HTG       08-JAN-2001
DEFINITION   Homo sapiens chromosome 14 clone R-728G21, *** SEQUENCING IN
             PROGRESS ***, in ordered pieces.
ACCESSION    AL390801
VERSION      AL390801.3  GI:12140328
KEYWORDS     HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 166847)
  AUTHORS    Genoscope.
  TITLE      Direct Submission
  JOURNAL    Submitted (08-JAN-2001) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
             - Web : www.genoscope.cns.fr)
             On Jan 12, 2001 this sequence version replaced gi:9716875.
COMMENT      ------------- Genome Center
             Center: Genoscope / Centre National de Sequencage
             Center code: GS
             Web site: http://www.genoscope.cns.fr/
             Contact: Seqref@genoscope.cns.fr
             -------------
             IMPORTANT: This sequence is unfinished and does not necessarily
             represent the correct sequence. Work on the sequence is in progress
             and the release of this data is based on the understanding that the
             sequence may change as work continue. The sequence may be
             contaminated with foreign sequence from E.coli, yeast, vector,
             phage, etc.. even if efforts are made to eliminate these
             contaminating sequences. The following BAC sequence is oriented
             from the T7 to the SP6 end.
             Upstream BAC (overlapping the T7 end) : R-629F19 (AC=AL391516)
             Downstream BAC (overlapping the SP6 end) : R-129M6 ------------
             Summary Statistics
             Assembly program: Phrap; version 2.0
             Quality coverage: 6.54x in Q20 bases; sum-of-contigs
             -------------
```

```
RESULT 6
CNS06C88/c
LOCUS       CNS06C88        167068 bp    DNA             HTG       08-DEC-2000
DEFINITION  Homo sapiens chromosome 14 clone R-629F19, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL391516
VERSION     AL391516.3  GI:11611190
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167068)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (07-DEC-2000) Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 9, 2000 this sequence version replaced gi:10046660.
            ----------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
```

```
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
--------
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc.. even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-728G21
Downstream BAC (overlapping the SP6 end) : R-757K12 (AC=AL390335)
              ---------- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 6.72x in Q20 bases; sum-of-contigs
----------

Overall quality chart :
Range    :    bases
0        :    12
1  - 9   :    30
10 - 19  :    60
20 - 29  :    217
30 - 39  :    1822
40 - 49  :    3286
50 - 59  :    6904
60 - 69  :    20683
70 - 79  :    55200
80 - 89  :    78854
90 - 99  :
----------
Percentage of bases with a quality value >= 40 : 99 %.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES             Location/Qualifiers
     source          1..167068
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-629F19"
                     /clone_lib="RPCI-11"
     STS             29922..30055
                     /note="matching EMBL:G30630
                     RHdb:RH37984
                     dbSTS:STS29655
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT      55505 a   29483 c   29391 g   52689 t
ORIGIN
```

```
Query Match         8.0%;     Score 35.6;   DB 84;   Length 167068;
Best Local Similarity  51.4%;    Pred. No. 4.8;
Matches 112;   Conservative  0;   Mismatches  99;   Indels  7;   Gaps  1;

Qy    209 tgttatcggagccattcgtgtgctgttactgtgtagcgcttgatgacttccttgctgta 268
          ||||||||||||||||  |||||||| || | || ||||||  | |||||||| | |||||
Db  85600 TGCCCTAGGAGGCCATTCACGACCAGGAATCTATTTAGCACTTGGTGACCCCGTGGTTGGT 85541

Qy    269 ctccgcttggagccattgagtgctgttaccgtgcagcactcgaagacttccttgctgt 328
          |||  |||  ||||||||| || |  |||||| ||||| |  |||
Db  85540 GTCCAGCCTCCTCCCCTTCAGTCCTTCAGCCCTGG------TGTCTGCCACCATCTTTCT 85488

Qy    329 actcgctggcagcatagtctgttctgccctggtcggggacaccctgtactacagtac 388
          |||  | |||  |  || |||| |||| |  ||| |||||  |||| ||
Db  85487 ATCCACCCTCAGTTGATTCTCTCAGAGGATCGTTTGGAAATGCTGGTGTACTTCATATG 85428

Qy    389 ttcgtttcccttgtagtggtactacttcttagccatg 426
          || | |||||||| ||||||  |||||
Db  85427 TTGTTCTCTTTGTGGGGAGAAGCTCTTCTTGGCCATG 85390
```

```
Overall quality chart :
Range    :    bases
0        :    212
1  - 9   :    7
10 - 19  :    54
20 - 29  :    134
30 - 39  :    428
40 - 49  :    2710
50 - 59  :    5039
60 - 69  :    6436
70 - 79  :    17421
80 - 89  :    48805
90 - 99  :    85601
----------
Percentage of bases with a quality value >= 40 : 99 %.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES             Location/Qualifiers
     source          1..166847
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-728G21"
                     /clone_lib="RPCI-11"
BASE COUNT      54079 a   28819 c   29091 g   54858 t
ORIGIN
```

```
Query Match         8.0%;     Score 35.6;   DB 84;   Length 166847;
Best Local Similarity  51.4%;    Pred. No. 4.8;
Matches 112;   Conservative  0;   Mismatches  99;   Indels  7;   Gaps  1;

Qy    209 tgttatcggagccattcgtgtgctgttactgtgtagcgcttgatgacttccttgctgta 268
          ||||||||||||||||  |||||||| || | || ||||||  | |||||||| | |||||
Db  65623 TGCCCTAGGAGGCCATTCACGACCAGGAATCTATTTAGCACTTGGTGACCCCGTGGTTGGT 65682

Qy    269 ctccgcttggagccattgagtgctgttaccgtgcagcactcgaagacttccttgctgt 328
          |||  |||  ||||||||| || |  |||||| ||||| |  |||
Db  65683 GTCCAGCCTCCTCCCCTTCAGTCCTTCAGCCCTGG------TGTCTGCCACCATCTTTCT 65735

Qy    329 actcgctggcagcatagtctgttctgccctggtcggggacaccctgtactacagtac 388
          |||  | |||  |  || |||| |||| |  ||| |||||  |||| ||
Db  65736 ATCCACCCTCAGTTGATTCTCTCAGAGGATCGTTTGGAAATGCTGGTGTACTTCATATG 65795

Qy    389 ttcgtttcccttgtagtggtactacttcttagccatg 426
          || | |||||||| ||||||  |||||
Db  65796 TTGTTCTCTTTGTGGGGAGAAGCTCTTCTTGGCCATG 65833
```

```
RESULT 7
CNS06C70/c
LOCUS       CNS06C70   176506 bp   DNA           HTG        02-AUG-2000
DEFINITION  Homo sapiens chromosome 14 clone R-757K12, *** SEQUENCING IN
            PROGRESS ***, in ordered pieces.
ACCESSION   AL390335
VERSION     AL390335.2  GI:9714131
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176506)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Aug 6, 2000 this sequence version replaced gi:9453789.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continue. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage, etc.. ..even if efforts are made to eliminate these
            contaminating sequences. The following BAC sequence is oriented
            from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-354G16
            Downstream BAC (overlapping the SP6 end) : R-629F19

            Overall quality chart :
            Range       :   bases
              0         :     237
              1 -   9   :      24
             10 -  19   :     133
             20 -  29   :     403
             30 -  39   :    1462
             40 -  49   :    7003
             50 -  59   :    8711
             60 -  69   :    7494
             70 -  79   :   17617
             80 -  89   :   46294
             90 -  99   :   87098

            Percentage of bases with a quality value >= 40 : 98 %.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..176506
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-757K12"
                     /clone_lib="RPCI-11"
BASE COUNT       56706 a   31526 c   31026 g   57248 t
ORIGIN

Query Match         8.0%;  Score 35.6;  DB 84;  Length 176506;
  Best Local Similarity   51.4%;
  Matches  112;  Conservative    0;  Mismatches   99;  Indels    7;  Gaps    1;

Qy    209 tgttatcggagccattcgtgtgctgttactgtgtagccgcttgatgacttcctgctgta 268
            | ||||||||||||   | ||| |||| | |||    |||| || ||||
Db  39886 TGCCCTAGGAGCCATTCACGACCAGGAATCTATTTAGCACTTGGTGACCCCGTGGTTGGT 39827

Qy    269 ctccgcttggagccattgagtgctgtttaccgtgcagcactcgaagacttcctgctgt 328
            |||   ||| |||| || ||| || |    |||| |  || || ||||
Db  39826 GTCCAGCCTCCTCCCCCTTCCCCCTTCAGCCCTGG------TGTCTGCCACCATCTTTCT 39774
```

```
Qy    329 actcgctggcagcatagtctgtttgctgcctgctcgtgtctctgggggacacctgtgtactacagtac 388
            | ||| |  ||||   |   ||| |  ||| |  |||| || || ||||||||  ||||||||
Db  39773 ATCCACCCTCAGTTGATTCTTCTCCAGAGGATCGTGTTTGGAAATGCTGGTGTACTTCATATG 39714

Qy    389 ttcgttccccttgtagtggtactacttcctttagccatg 426
            ||||||||||| || ||  ||| |||||| ||||
Db  39713 TTGTTCTCTCTTGTGGGGAGAAGCTCTTCTTGGCCATG 39676
```

```
RESULT 8
BCHMLYSN
LOCUS       BCHMLYSN        1042 bp    DNA          BCT       17-NOV-1995
DEFINITION  B.cereus gene for hemolytic factor.
ACCESSION   X84058
VERSION     X84058.1  GI:662879
KEYWORDS    hemolysin; hemolytic factor.
SOURCE      Bacillus cereus.
  ORGANISM  Bacillus cereus
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.
REFERENCE   1  (bases 1 to 1042)
  AUTHORS   Baida,G.E. and Kuzmin,N.P.
  TITLE     Cloning and primary structure of a new hemolysin gene from Bacillus
            cereus
  JOURNAL   Biochim. Biophys. Acta 1264 (2), 151-154 (1995)
  MEDLINE   96085115
REFERENCE   2  (bases 1 to 1042)
  AUTHORS   Baida,G.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-1995) Baida G. E., Institute of Biochemistry and
            Physiology of Microorganisms, Laboratory of Gene Systematics, Pr.
            Nauki 5, Pushchino, Moscow region, Russian Federation, 142292
FEATURES             Location/Qualifiers
     source          1..1042
                     /organism="Bacillus cereus"
                     /strain="VKM-B164"
                     /db_xref="taxon:1396"
     repeat_unit     172..184
                     /note="2nd location 213. . .225; contains putative
                     -35_signal TTGTCA"
                     /rpt_type=DIRECT
     RBS             321..335
                     /note="includes putative SD-sequence and initiation ATG
                     codon"
     CDS             333..992
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="novel hemolytic factor"
                     /protein_id="CAA58877.1"
                     /db_xref="GI:662880"
                     /db_xref="SWISS-PROT:P54176"
                     /translation="MTEKMTRMTQFVKEEIANAITHGIGAILSIPALIILIIHASKHG
                     TASAVVAFTVYGVSMFLLYLFSTLLHSIIHHPKVEKLFTILDHSAIYLLIAGTYPFLL
                     ITLRGPLGWTLLAIIWTLAIGGIIFKIFFVRRFIKASTLCYIIMGWLIIVAIKPLYEN
                     LTGHGFSLLLAGGILYSVGAIFFLWEKLPFNHAIWHLFVLGGSAMMFFCVLFYVLPTA
                     S"
     stem_loop       989..1031
                     /note="supposed transcription terminator"
BASE COUNT         319 a    191 c    157 g    375 t
ORIGIN

Query Match         7.9%;  Score 35.2;  DB 2;  Length 1042;
  Best Local Similarity   54.7%;  Pred. No. 5.8;
  Matches   70;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    208 atgttatcggagccattcgtgtgctgttactgtgtagcgcttgatgacttcctgctgt 267
            ||||| | ||| ||||| || ||  |  ||| ||| |||  ||||| ||||||||
Db    460 ATGGTACCGCATCTGCTGTTGTTGCATTACGGTTACCGTTGAAGCGTGTAAGCATGTTCTTACTGT 519

Qy    268 actccgcttggagccattgagtgctgtttaccgtgcagcactcgagacttcctgctg 327
            ||| |  ||  || || |  ||| ||| | ||||   || ||| |||||||||||
```

```
Db      520 ACTTGTTTCGACATTGTTACATAGCATTCACCATTCCAAAAGTAGAAAAATTATTACTA 579
Qy      328 tactcgct 335
            |||| ||
Db      580 TACTAGAT 587
```

RESULT 9
```
LOCUS       I66494                  7218 bp    DNA       PAT       28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1..7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN
```

```
Query Match             7.8%;  Score 34.8;  DB 10;  Length 7218;
  Best Local Similarity   3.7%;
  Matches   12;  Conservative  175;  Mismatches  137;  Indels    0;  Gaps    0;

Qy      117 tccctgcccttgagggtcgtctcaaacatgcccctcgagacgtcggtgaagccgtagt 176
                  ::::: :::::::::::::::::::::::::::::::::::::::::::::::
Db     1133 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1192
Qy      177 ttgtacaagactaacaacgccaatggcgtcgatgttatcggagccattcgtgctgcgtt 236
                ::: ::::::: :::::::::::::::::::::::::::::::::::::::::::::
Db     1193 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1252
Qy      237 actgtgtagcgcttgatgacttcctgctgctgtactccgcttggagccattgagtgctgtt 296
                ::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db     1253 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1312
Qy      297 taccgtgcagcagcactcgaagacttcctgctgctcgctggcagcagcatagctgttgctg 356
                ::: :::::: :::::::::::::::::::::::::::::::::::::::::::
Db     1313 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1372
Qy      357 cctggtctgggacacctgtgtactacagtactccgtttccctttgtagtggtactactt 416
                :::: :::::: ::::::::::::::::::::::::::::::::::::::::
Db     1373 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1432
Qy      417 cttagccatgtactcgtgtctgtt 440
                ||| ||::||:||:||||||
Db     1433 YYGTACCAAATTCTTCTATCTCTT 1456
```

RESULT 10
AC073824/c
```
LOCUS       AC073824             180944 bp    DNA       HTG       29-JUN-2000
DEFINITION  Mus musculus clone RP23-92G22, WORKING DRAFT SEQUENCE, 48 unordered
            pieces.
ACCESSION   AC073824
VERSION     AC073824.1  GI:8810441
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 180944)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 180944)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov

            Project Information
            Center Project Name: 1772366
            Center clone name: RPCI-23_92G22
            -----
            Summary Statistics
            Consensus quality: 143665 bases at least Q40
            Consensus quality: 160562 bases at least Q30
            Consensus quality: 164105 bases at least Q20
            Estimated insert size: 194000: pulse field gel estimation
            Estimated insert size: 176244; sum-of-contigs estimation
            Quality coverage:   6.92 in Q20 bases; pulse field gel estimation
            Quality/Qualifiers  7.62 in Q20 bases; sum-of-contigs estimation
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 48 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *     1102:     contig of 1102 bp in length
          *     1103:
          *     1202:     gap of unknown length
          *     1203:
          *     2608:     contig of 1406 bp in length
          *     2609:
          *     2708:     gap of unknown length
          *     2709:
          *     3827:     contig of 1119 bp in length
          *     3828:
          *     3927:     gap of unknown length
          *     3928:
          *     5266:     contig of 1339 bp in length
          *     5267:
          *     5366:     gap of unknown length
          *     5367:
          *     6856:     contig of 1490 bp in length
          *     6857:
          *     6956:     gap of unknown length
          *     6957:
          *     8006:     contig of 1050 bp in length
          *     8007:
          *     8106:     gap of unknown length
          *     8107:
          *     9357:     contig of 1251 bp in length
          *     9358:
          *     9457:     gap of unknown length
          *     9458:
          *    10998:     contig of 1541 bp in length
          *    10999:
          *    11098:     gap of unknown length
          *    11099:
          *    12370:     contig of 1272 bp in length
          *    12371:
          *    12470:     gap of unknown length
          *    12471:
          *    13567:     contig of 1097 bp in length
          *    13568:
          *    13667:     gap of unknown length
          *    13668:
          *    15007:     contig of 1340 bp in length
          *    15008:
          *    15107:     gap of unknown length
          *    15108:
          *    17117:     contig of 2010 bp in length
          *    17118:
          *    17217:     gap of unknown length
          *    17218:
          *    18346:     contig of 1129 bp in length
          *    18347:
          *    18446:     gap of unknown length
          *    18447:
          *    19824:     contig of 1378 bp in length
          *    19825:
          *    19924:     gap of unknown length
          *    19925:
          *    21617:     contig of 1693 bp in length
          *    21618:
          *    21717:     gap of unknown length
          *    21718:
          *    23913:     contig of 2196 bp in length
          *    23914:
          *    24013:     gap of unknown length
          *    24014:
          *    25142:     contig of 1129 bp in length
          *    25143:
          *    25242:     gap of unknown length
          *    25243:
          *    26540:     contig of 1298 bp in length
          *    26541:
          *    26640:     gap of unknown length
          *    26641:
          *    28457:     contig of 1817 bp in length
          *    28458:
          *    28557:     gap of unknown length
          *    28558:
          *    30203:     contig of 1646 bp in length
          *    30204:
          *    30303:     gap of unknown length
          *    30304:
          *    31545:     contig of 1242 bp in length
          *    31546:
          *    31645:     gap of unknown length
          *    31646:
          *    32767:     contig of 1122 bp in length
```

```
RESULT 11
AL133174/c
LOCUS       AL133174        113836 bp    DNA                          PRI    06-MAR-2001
DEFINITION  Human DNA sequence from clone RP3-470L14 on chromosome 20 Contains
            the CSEIL gene for chromosome segregation 1 (yeast homolog)-like
            protein, the STAU gene for Staufen (RNA-binding protein), a novel
            gene similar to ARC21 encoding the Arp2/3 protein complex subunit
            p21-Arc, a CpG island, ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL133174
VERSION     AL133174.15   GI:8573761
KEYWORDS    HTG; ARC21; Arp2/3; chromosome segregation; CSEIL; p21-Arc;
            RNA-binding protein; STAU; staufen.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113836)
  AUTHORS   Ramsay,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 20, 2000 this sequence version replaced gi:8388434.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            IMPORTANT: This sequence is not the entire insert of clone
            RP3-470L14 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP3-470L14 is at 113836 in this
            sequence. The true right end of clone RP1-155G6 is at 100 in this
            sequence. This sequence was finished as follows unless otherwise
            noted: all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            RP3-470L14 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
VECTOR:     pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..113836
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /clone="RP3-470L14"
                     /clone_lib="RPCI-3"
     repeat_region   1..139
                     /note="AluSg/x repeat: matches 136. .274 of consensus"
     repeat_region   140..161
                     /note="TIGGER2 repeat: matches 6. .27 of consensus"
     repeat_region   316..449
                     /note="AluY repeat: matches 136. .229 of consensus"
     repeat_region   558. .847
                     /note="AluSp repeat: matches 1. .299 of consensus"
     repeat_region   855. .1030
                     /note="AluJo repeat: matches 120. .291 of consensus"
```

```
*    32768    32867:  gap of unknown length
*    32868    34906:  contig of 2039 bp in length
*    34907    35006:  gap of unknown length
*    35007    39922:  contig of 4916 bp in length
*    39923    40022:  gap of unknown length
*    40023    43132:  contig of 3110 bp in length
*    43133    43232:  gap of unknown length
*    43233    45991:  contig of 2759 bp in length
*    45992    46091:  gap of unknown length
*    46092    48619:  contig of 2528 bp in length
*    48620    48719:  gap of unknown length
*    48719    53180:  contig of 4461 bp in length
*    53181    53280:  gap of unknown length
*    53281    55680:  contig of 2400 bp in length
*    55681    55780:  gap of unknown length
*    55781    60954:  contig of 5174 bp in length
*    60955    61054:  gap of unknown length
*    61055    65155:  contig of 4101 bp in length
*    65156    65255:  gap of unknown length
*    65256    69032:  contig of 3777 bp in length
*    69033    69132:  gap of unknown length
*    69133    71751:  contig of 2619 bp in length
*    71752    71851:  gap of unknown length
*    71852    77288:  contig of 5437 bp in length
*    77289    77388:  gap of unknown length
*    77389    85634:  contig of 8246 bp in length
*    85635    85734:  gap of unknown length
*    85735    92244:  contig of 6510 bp in length
*    92245    92344:  gap of unknown length
*    92345    98717:  contig of 6373 bp in length
*    98718    98817:  gap of unknown length
*    98818   101003:  contig of 2186 bp in length
*   101004   101103:  gap of unknown length
*   101104   106020:  contig of 4917 bp in length
*   106021   106120:  gap of unknown length
*   106121   111507:  contig of 5387 bp in length
*   111508   111607:  gap of unknown length
*   111608   119749:  contig of 8142 bp in length
*   119750   119849:  gap of unknown length
*   119849   127441:  contig of 7592 bp in length
*   127442   127541:  gap of unknown length
*   127542   136208:  contig of 8667 bp in length
*   136209   136308:  gap of unknown length
*   136309   145027:  contig of 8719 bp in length
*   145028   145127:  gap of unknown length
*   145128   154394:  contig of 9267 bp in length
*   154395   154494:  gap of unknown length
*   154495   163234:  contig of 8740 bp in length
*   163235   163334:  gap of unknown length
*   163335   170038:  contig of 6704 bp in length
*   170039   170138:  gap of unknown length
*   170139   180944:  contig of 10806 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .180944
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-92G22"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    49088 a  39940 c  40162 g  47035 t  4719 others
ORIGIN
```

```
Query Match          7.8%;    Score 34.8;    DB 75;    Length 180944;
Best Local Similarity 58.8%;  Pred. No. 8.5;
Matches   60;  Conservative  0;  Mismatches  42;  Indels  0;  Gaps  0;

Qy    318 tcccttgctgtactcgctggcagcatagtctgtttgcctggtctgggggacacctgtg 377
          ||||||| || ||  ||||| | ||  |||| |||| |||| ||||||| ||||| |||
Db   6764 TTGGTTTCGCTATTCTCGGGCCTCCTCGTCGTTTGGCTCGTCGTTCGTGCGTGCTTTGC 6705

Qy    378 tactacagtactcgttccccttgtagtggtactacttctt 419
          | |||| ||||  ||||||| ||  |||  |||   |||||
Db   6704 TGGTTCTGTCGTTGTTGTTTCGCTTGTTTTGTTCCTCTTGTTTCTT 6663
```

```
misc_feature      1044. .1352
                  /note="match: STS: Em:G42844"
repeat_region     1044. .1357
repeat_region     1360. .1519
                  /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region     1586. .1760
                  /note="FRAM repeat: matches 1. .160 of consensus"
misc_feature      1723. .2467
                  /note="LIM1 repeat: matches 5299. .5484 of consensus"
misc_feature      complement(2124. .2394)
                  /note="match: GSS: Em:AQ896083"
misc_feature      complement(2124. .2394)
                  /note="match: GSS: Em:AQ313063"
misc_feature      2135. .2387
                  /note="match: GSS: Em:AQ200628"
misc_feature      2136. .2430
                  /note="match: GSS: Em:AQ351010"
misc_feature      2141. .2370
                  /note="match: GSS: Em:AQ318962"
misc_feature      complement(2142. .2532)
                  /note="match: GSS: Em:AQ314242"
misc_feature      complement(2143. .2448)
                  /note="match: GSS: Em:AQ028497"
misc_feature      complement(2149. .2519)
                  /note="match: GSS: Em:AF010903"
misc_feature      complement(2150. .2517)
                  /note="match: GSS: Em:AQ341931"
misc_feature      complement(2155. .2387)
                  /note="match: STS: Em:G29081"
misc_feature      complement(2155. .2792)
                  /note="match: GSS: Em:AQ357800"
misc_feature      complement(2161. .2428)
                  /note="match: GSS: Em:AQ764595"
misc_feature      complement(2161. .2516)
                  /note="match: GSS: Em:AQ532045"
misc_feature      complement(2161. .2470)
                  /note="match: GSS: Em:AQ169318"
misc_feature      complement(2161. .2535)
                  /note="match: GSS: Em:B84560"
misc_feature      2162. .2394
                  /note="match: GSS: Em:AQ413395"
misc_feature      2162. .2517
                  /note="match: GSS: Em:AQ270015"
misc_feature      complement(2162. .2456)
                  /note="match: GSS: Em:AQ060045"
misc_feature      2162. .2456
                  /note="match: GSS: Em:B51613"
misc_feature      complement(2167. .2387)
                  /note="match: GSS: Em:AQ281464"
misc_feature      complement(2167. .2526)
                  /note="match: GSS: Em:AQ497805"
misc_feature      complement(2167. .2524)
                  /note="match: GSS: Em:AQ497783"
misc_feature      complement(2167. .2531)
                  /note="match: GSS: Em:AQ058261"
misc_feature      2167. .2515
                  /note="match: GSS: Em:AQ269024"
misc_feature      complement(join(2168. .2204,2642. .2882))
                  /note="match: GSS: Em:AQ749070"
misc_feature      complement(2168. .2504)
                  /note="match: GSS: Em:AQ535341"
misc_feature      complement(2278. .2534)
                  /note="match: GSS: Em:AG012200"
misc_feature      2301. .2510
                  /note="match: STS: Em:G10526"
misc_feature      complement(2306. .2545)
                  /note="match: GSS: Em:AQ055853"
misc_feature      complement(2306. .2535)
                  /note="match: GSS: Em:B91074"
misc_feature      complement(2306. .2519)
                  /note="match: GSS: Em:AQ240974"
repeat_region     2314. .2515
                  /note="AluSq repeat: matches 1. .210 of consensus"
misc_feature      2314. .2504

misc_feature      complement(2318. .2513)
                  /note="match: STS: Em:HSC84B1"
misc_feature      2318. .2516
                  /note="match: GSS: Em:AQ138787"
misc_feature      2389. .2517
                  /note="match: GSS: Em:AQ013658"
misc_feature      2391. .2526
                  /note="match: GSS: Em:AQ320309"
misc_feature      2391. .2534
                  /note="match: STS: Em:G37384"
misc_feature      2391. .2514
                  /note="match: GSS: Em:AQ741808"
misc_feature      2391. .2516
                  /note="match: GSS: Em:AQ894888"
repeat_region     2565. .2892
                  /note="AluY repeat: matches 1. .311 of consensus"
misc_feature      2642. .2918
                  /note="match: GSS: Em:AQ788416"
misc_feature      2642. .2885
                  /note="match: GSS: Em:AQ744233"
repeat_region     2893. .3052
                  /note="FRAM repeat: matches 4. .162 of consensus"
misc_feature      3061. .3077
                  /note="Weak data"
repeat_region     3085. .3393
                  /note="AluY repeat: matches 1. .311 of consensus"
repeat_region     3425. .3638
                  /note="AluJo repeat: matches 3. .204 of consensus"
mRNA              join(<4098. .4234,5355. .5458,8081. .8178,9319. .9469,
                  9610. .9818,10814. .10911,11009. .11094,12118. .12199,
                  14212. .14358,14804. .15035,16421. .17024)
                  /gene="CSE1L"
                  /note="match: cDNAs: Em:U33286 Em:AF053641 Em:AB036757"
                  /evidence=not_experimental
                  /product="dJ470L14.1.1 (isoform 1 of chromosome
                  segregation 1 (yeast homolog)-like (CSE1L) )"
                  4098. .17024
gene              4098. .17024
                  /gene="CSE1L"
CDS               join(<4098. .4234,5355. .5458,8081. .8178,9319. .9469,
                  9610. .9818,10814. .10911,11009. .11094,12118. .12199,
                  14212. .14358,14804. .15035,16421. .16510)
                  /gene="CSE1L"
                  /note="continues in clone dJ155G6 (Em:AL121903) as
                  dJ155G6.2.1
                  match: proteins: Sw:P55060 Tr:Q9PTU3 Tr:Q9XZU1"
                  /codon_start=1
                  /evidence=not_experimental
                  /product="dJ470L14.1.1 (isoform 1 of chromosome
                  segregation 1 (yeast homolog)-like (CSE1L) )"
                  /protein_id="CAC14081.1"
                  /db_xref="GI:10944131"
                  /translation="VPKEHLLVSIPLLINHLQAESIVHTYAAHALERLFTMRGPNNA
TLFTAABIAPFVEILLTNLFKALFLPGSSENEVIMKAIMRSFSLLQEAIIPYIPTLIT
QLTQKLLAVSKNPSKPHFNHYMFEAICLSIRITCKANPAAVNFEEALFLVFTEILQN
DVQEFIPYVFQVMSLLETHKNDIPSSYMALFPHLLQPVLWERTGNIPALVRLLQAFL
ERGSNTIASAADKIPGLLGVFQKLIASKANDHQGFYILLNSIEHMPPESVDQYRKQI
FILLFQRLQNSKTTKFIKSFLVFINLYCIKYGALALQEIFDGIQPKMFGMVLEKIIIP
```

```
Query Match           7.8%;  Score 34.6;  DB 89;  Length 113836;
Best Local Similarity 58.1%;  Pred. No. 9.7;
Matches    61;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    166 gaagccgtagtttgtacaagactaacaacgccaatggcgtcagtgttatcggagccattc 225
             | |||||   || ||  || ||  | | |||  | | || |||  ||||||||||||
Db  25988 GAGGTGGGAGTTGTTAATGGCAGCATAGCCACCTTCTTGGATTTTTGTAAACCTTTC 25929

Qy    226 gtgtgctgttactgtagcgcttgatgacttccttgctgtact 270
             |||| |||||||| |  ||  ||||| |||| ||| |||||||
Db  25928 CTGTGGGTGCTGAGTTCGTGGAACTTGTTACTTTGTGTACTGTACT 25884
```

RESULT   12

```
AF123274/c
LOCUS       AF123274       397 bp    DNA            MAM       27-MAY-1999
DEFINITION  Physeter catodon strain macrocephalus microsatellite JB69 sequence.
ACCESSION   AF123274
VERSION     AF123274.1  GI:4894613
KEYWORDS    .
SOURCE      sperm whale.
ORGANISM    Physeter catodon
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
            Physeteridae; Physeter.
REFERENCE   1  (bases 1 to 397)
AUTHORS     Bond,J.M. and Amos,B.
TITLE       Sperm whale microsatellite loci
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 397)
AUTHORS     Bond,J.M. and Amos,B.
TITLE       Direct Submission
JOURNAL     Submitted (22-JAN-1999) Molecular Ecology Group, Zoology, Cambridge
            University, Downing Street, Cambridge CB2 3EJ, UK
FEATURES             Location/Qualifiers
     source          1..397
                     /organism="Physeter catodon"
                     /db_xref="taxon:9755"
                     /sex="male"
                     /tissue_type="skin"
                     /dev_stage="immature"
                     /note="Physeter macrocephalus
                     primer A: tagatcctagaggtaacactgaac; primer B:
                     ctcctacctaggtactctgtact; PCR conditions: 2 mins @94
                     degrees x1, 1 min @94 degrees, 30 secs @47 degrees, 50
                     secs @72 degrees x5, 45 secs @90 degrees, 1 min @51
                     degrees, 1 min @72 degrees x25; reaction conditions: 1.5
                     mM MgCl2, 60 mM TMAC, 2.5% formamide"
     repeat_region   1..397
                     /note="microsatellite JB69; 20 alleles identified in 274
                     whales"
                     /rpt_type=tandem
BASE COUNT      104 a      77 c     103 g    112 t      1 others
ORIGIN

Query Match         7.7%;       Score 34.2;   DB 7;   Length 397;
Best Local Similarity   51.3%;   Pred. No. 12;
Matches   78;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy     14 ctatgtgatgcaggagggcctagtggtcatctgaaggaaaggacaaagacacctccaag 73
          |||  | ||   ||||| | |||||| || |||| || |||| || |||| |||||||
Db    357 CTTGCCTGGGTGAAAAGTTCAGTGGTACAAGAGCAGNCAATGAAAAGTCATCTCCCTC 298

Qy     74 ccggacacggtactgtagctgcacggcataatggaaaacagtccctgccctttgagg 133
          ||  | || ||| || ||| | ||| ||||||||| || ||| |  |||||| |||||
Db    297 CCAGTACAGAGTACCCCAGGTAGGAGCACCCCAGAGATACCCGTTGTCGATTTCTTGGGG 238

Qy    134 gtcgtctcaaacatgccctcgagacgtcggt 165
          ||  | |||||||  | ||||| |  |||||
Db    237 ATCCTTCCAGACATGCATTTACAGTCTTTTGT 206
```

RESULT 13

```
AC068473
LOCUS       AC068473     193583 bp    DNA              HTG      18-MAR-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-567M16 map 18, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC068473
VERSION     AC068473.4  GI:13376972
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193583)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
                                                                      TITLE      Homo sapiens chromosome 18, clone RP11-567M16
                                                                      JOURNAL    Unpublished
                                                                      REFERENCE  2  (bases 1 to 193583)
                                                                      AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
                                                                                 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
                                                                                 Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
                                                                                 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
                                                                                 Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
                                                                                 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
                                                                                 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
                                                                                 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
                                                                                 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
                                                                                 Klein,J., LaRocque,R., Lamazares,R., Landers,T., Lehoczky,J.,
                                                                                 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
                                                                                 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
                                                                                 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
                                                                                 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
                                                                                 O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
                                                                                 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
                                                                                 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
                                                                                 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
                                                                                 Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
                                                                                 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
                                                                                 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
                                                                      TITLE      Direct Submission
                                                                      JOURNAL    Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
                                                                                 Research, 320 Charles Street, Cambridge, MA 02141, USA
                                                                      COMMENT    On Mar 18, 2001 this sequence version replaced gi:13357529.
                                                                                 All repeats were identified using RepeatMasker.
                                                                                 Smit, A.F.A. & Green, P. (1996-1997)
                                                                                 http://ftp.genome.washington.edu/RM/RepeatMasker.html
                                                                                 --------- Genome Center
                                                                                 Center: Whitehead Institute/ MIT Center for Genome Research
                                                                                 Center code: WIBR
                                                                                 Web site: http://www-seq.wi.mit.edu
                                                                                 Contact: sequence_submissions@genome.wi.mit.edu
                                                                                 --------- Project Information
                                                                                 Center project name: L9755
                                                                                 Center clone name: 567_M16
                                                                                 --------- Summary Statistics
                                                                                 Sequencing vector: M13; M77815; 47% of reads
                                                                                 Sequencing vector: Plasmid; n/a; 53% of reads
                                                                                 Chemistry: Dye-terminator Big Dye; 100% of reads
                                                                                 Assembly program: Phrap; version 0.960731
                                                                                 Consensus quality: 188595 bases at least Q40
                                                                                 Consensus quality: 190640 bases at least Q30
                                                                                 Consensus quality: 191543 bases at least Q20
                                                                                 Insert size: 167000; agarose-fp
                                                                                 Insert size: 192483; sum-of-contigs
                                                                                 Quality coverage: 8.6 in Q20 bases; agarose-fp
                                                                                 Quality coverage: 7.4 in Q20 b.
                                                                      * NOTE: This is a 'working draft' sequence. It currently
                                                                      * consists of 12 contigs. The true order of the pieces
                                                                      * is not known and their order in this sequence record is
                                                                      * arbitrary. Gaps between the contigs are represented as
                                                                      * runs of N, but the exact sizes of the gaps are unknown.
                                                                      * This record will be updated with the finished sequence
                                                                      * as soon as it is available and the accession number will
                                                                      * be preserved.
                                                                      *        1    2398: contig of 2398 bp in length
                                                                      *     2399  2498: gap of    100 bp
                                                                      *     2499  3047: contig of 549 bp in length
                                                                      *     3048  3147: gap of    100 bp
                                                                      *     3148  3394: contig of 247 bp in length
                                                                      *     3395  3494: gap of    100 bp
                                                                      *     3495 33493: contig of 29999 bp in length
                                                                      *    33494 33593: gap of    100 bp
                                                                      *    33594 42318: contig of 8725 bp in length
                                                                      *    42319 42418: gap of    100 bp
                                                                      *    42419 51882: contig of 9464 bp in length
                                                                      *    51883 51982: gap of    100 bp
                                                                      *    51983 67955: contig of 15973 bp in length
                                                                      *    67956 68055: gap of    100 bp
```

```
                 *   68056    83830: contig of 15775 bp in length
                 *   83831    83930: gap of    100 bp
                 *   83931   104846: contig of 20916 bp in length
                 *  104847   104946: gap of    100 bp
                 *  104947   136507: contig of 31561 bp in length
                 *  136508   136607: gap of    100 bp
                 *  136608   161450: contig of 24843 bp in length
                 *  161451   161550: gap of    100 bp
                 *  161551   193583: contig of 32033 bp in length.
DEFINITION   Homo sapiens clone RP11-9D19, WORKING DRAFT SEQUENCE, 11 unordered
             pieces.
ACCESSION    AC021765
VERSION      AC021765.3  GI:10198414
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 152393)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens, clone RP11-9D19
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 152393)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
             Boguslavily,L., Boukhgaiter,B., Brown,A., Burkett,G., Castle,A.,
             Choepel,Y., Coiangelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
             Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
             Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
             Norman,C.H., O'Connor,T.M., O'Donnell,P., Olivar,T.M., Peterson,K.,
             Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
             Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
             Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
             Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Sep 20, 2000 this sequence version replaced gi:9123795.
             All repeats were identified using RepeatMasker.
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ----------------- Genome Center
                 Center: Whitehead Institute/ MIT Center for Genome Research
                 Center code: WIBR
                 Web site: http://www-seq.wi.mit.edu
                 Contact: sequence_submissions@genome.wi.mit.edu
             ----------------- Project Information
                 Center project name: L3002
                 Center clone name: 9_D_19
             ----------------- Summary Statistics
                 Sequencing vector: M13; M77815; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.960731
                 Consensus quality: 146933 bases at least Q40
                 Consensus quality: 149520 bases at least Q30
                 Consensus quality: 150601 bases at least Q20
                 Insert size: 157000; agarose-fp
                 Insert size: 151393; sum-of-contigs
                 Quality coverage: 4.6 in Q20 bases; agarose-fp
                 Quality coverage: 4.7 in Q20 bases; sum-of-contigs
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 11 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1    28910: contig of 28910 bp in length
             *   28911    29010: gap of    100 bp
             *   29011    31437: contig of 2427 bp in length
             *   31438    31537: gap of    100 bp
             *   31538    33952: contig of 2415 bp in length
             *   33953    34052: gap of    100 bp
             *   34053    38372: contig of 4320 bp in length
FEATURES             Location/Qualifiers
     source          1..193583
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-567M16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..2398
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    2499..3047
                     /note="assembly_fragment"
     misc_feature    3148..3394
                     /note="assembly_fragment"
     misc_feature    3495..33493
                     /note="assembly_fragment"
     misc_feature    33594..42318
                     /note="assembly_fragment"
     misc_feature    42419..51882
                     /note="assembly_fragment"
     misc_feature    51983..67955
                     /note="assembly_fragment"
     misc_feature    68056..83830
                     /note="assembly_fragment"
     misc_feature    83931..104846
                     /note="assembly_fragment"
     misc_feature    104947..136507
                     /note="assembly_fragment"
     misc_feature    136608..161450
                     /note="assembly_fragment"
     misc_feature    161551..193583
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    45598 a  49135 c  50190 g  47399 t   1261 others
ORIGIN

Query Match       7.7%;    Score 34.2;    DB 73;    Length 193583;
Best Local Similarity   47.9%;    Pred. No. 13;
Matches   93;  Conservative   0;  Mismatches  101;  Indels   0;  Gaps   0;

Qy    85 tactgtagctggcacggcataatggaaacagtccctgccctttgagggtcgtcctcaaa 144
         ||| |||||| |||  | |-||-|-|-|| | ||| || ||-| | |||| |-|| |||
Db 20444 TAGTGACTCTGGCGCTTAAGAAACAGTTTGGGAACTCCTGCCTGCGAAGGAGAATCGGAAA 20503

Qy   145 catgccccctgagacgtcggtgaagccgtagtttgtacaagagactaacacgccaatggcg 204
         || | |||| ||-|||| |-|||| || |-|-|| ||| -|-|  | |-| |||| ||||
Db 20504 CANGCACTGNAAGATGTTACTGAGGTTAAATGCGACCCAAACAGAAAAGGGNCTTCGGGG 20563

Qy   205 tcgatgttatcggagccatccgtcgctgtcggttactgtgtagcgctctgatgactcctgc 264
         || ||| ||||| ||||| |||-|||| |||-|||| ||||| |-||| |-|-|||-|||
Db 20564 CGGTTGGTGTCCGTCTGATTAACCTGCTGTCGTCTGCTGGTTACATGTGGATTAGAATCTTAA 20623

Qy   265 tgtactccgctgg 278
         | ||||||||||
Db 20624 ACTGTTACGCTTGG 20637

RESULT 14
AC021765
LOCUS        AC021765    152393 bp    DNA            HTG        20-SEP-2000
```

```
VERSION     AC011159.3  GI:7107892
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158113)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-9D22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158113)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,J., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6539356.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
              Center project name: L3003
              Center clone name: 9_D_22
            --------------- Summary Statistics
              Sequencing vector: M13; M77815; 100% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 139493 bases at least Q40
              Consensus quality: 149656 bases at least Q30
              Consensus quality: 154385 bases at least Q20
              Insert size: 157000; agarose-fp
              Insert size: 157513; sum-of-contigs
              Quality coverage: 5.5 in Q20 bases; agarose-fp
              Quality coverage: 5.5 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1   1274: contig of 1274 bp in length
            *   1275   1374: gap of    100 bp
            *   1375   2530: contig of 1156 bp in length
            *   2531   2630: gap of    100 bp
            *   2631   3957: contig of 1327 bp in length
            *   3958   4057: gap of    100 bp
            *   4058  15747: contig of 11690 bp in length
            *  15748  15847: gap of    100 bp
            *  15848  49213: contig of 33366 bp in length
            *  49214  49313: gap of    100 bp
            *  49314  91982: contig of 42669 bp in length
            *  91983  92082: gap of    100 bp
            *  92083 158113: contig of 66031 bp in length.
```

```
*    38373  38472: gap of       100 bp
*    38473  81141: contig of 42669 bp in length
*    81142  81241: gap of       100 bp
*    81242  86894: contig of 5653 bp in length
*    86895  86994: gap of       100 bp
*    86995  95082: contig of 8088 bp in length
*    95083  95182: gap of       100 bp
*    95183 110046: contig of 14864 bp in length
*   110047 110146: gap of       100 bp
*   110147 129817: contig of 19671 bp in length
*   129818 129917: gap of       100 bp
*   129918 148772: contig of 18855 bp in length
*   148773 148872: gap of       100 bp
*   148873 152393: contig of 3521 bp in length.
FEATURES             Location/Qualifiers
     source          1..152393
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-9D19"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..28910
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    29011..31437
                     /note="assembly_fragment"
     misc_feature    31538..33952
                     /note="assembly_fragment"
     misc_feature    34053..38372
                     /note="assembly_fragment"
     misc_feature    38473..81141
                     /note="assembly_fragment"
     misc_feature    81242..86894
                     /note="assembly_fragment"
     misc_feature    86995..95082
                     /note="assembly_fragment"
     misc_feature    95183..110046
                     /note="assembly_fragment"
     misc_feature    110147..129817
                     /note="assembly_fragment"
     misc_feature    129918..148772
                     /note="assembly_fragment"
     misc_feature    148873..152393
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49730 a  27512 c  26497 g  47643 t    1011 others
ORIGIN
```

```
Query Match        7.7%;    Score 34;   DB 66;  Length 152393;
Best Local Similarity  53.8%;   Pred No. 15;
Matches   70; Conservative   0; Mismatches   60; Indels   0; Gaps   0;

Qy   107 tgaaaacagtccctgcccttttgagggtcgtctcaaacatgcccctcgagacgtcggtg 166
             ||  ||  ||||||| |  ||||| | |  |  ||| |||| ||| || |||  ||||
Db  2644 TGGGCAAGTTTTCAAGCCCTTCCATCTTCCACCTGGAAATGGACTCAGAGATGTTGGTG 2703

Qy   167 aagccgtagtttgtacaagactaacaacgccaatggcgtcgatgttatcggagccattcg 226
            |  ||| ||||  | ||||| || |  ||  | ||| | | |  ||| || | ||
Db  2704 GGGGCACAGTGGGAATAAGACTGACTCTTTGGTTGCATGGATGCTGGGTGAGGCCCTGTG 2763

Qy   227 tgtgctgtt 236
            ||||| ||
Db  2764 ATTGCTGGTT 2773
```

```
RESULT 15
AC011159/c
LOCUS       AC011159   158113 bp    DNA         HTG       12-MAR-2000
DEFINITION  Homo sapiens clone RP11-9D22, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   AC011159
```

```
FEATURES
source          Location/Qualifiers
                1..158113
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="RP11-9D22"
                /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1274
                /note="assembly_fragment"
     misc_feature    1375..2530
                /note="assembly_fragment"
     misc_feature    2631..3957
                /note="assembly_fragment"
     misc_feature    4058..15747
                /note="assembly_fragment"
     misc_feature    15848..49213
                /note="assembly_fragment"
     misc_feature    49314..91982
                /note="assembly_fragment"
                clone_end:T7
                vector_side:right"
     misc_feature    92083..158113
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:right"
BASE COUNT  50587 a  26938 c  29440 g  50534 t   614 others
ORIGIN

Query Match             7.7%;   Score 34;  DB 62;  Length 158113;
Best Local Similarity   53.8%;  Pred. No. 15;
Matches  70;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    107 tggaaaacagtccctgcccttttgagggtcgtctcaaacatgcccctcgagacgtcggtg 166
          ||||| || |||||||| | |||||| |||||||||| |||||||| | || |||||||
Db 155470 TGGGCAAGTTTTCAAGCCCTTCCATCTTCCACCTGGAAATGGACTCAGAGATGTTGGTG 155411

Qy    167 aagccgtagtttgtacaagactaacacgccaatggcgtcgatgttatcggagccattcg 226
          | ||| | ||| |||||| |||| |||| || ||||||  | ||| ||| ||||||
Db 155410 GGGGCACAGTGGGAATAAGACTGACTCTTTGGTTTGCATGGATGCTGGGTGAGGCCTGTG 155351

Qy    227 tgtgctgttt 236
          ||||| ||
Db 155350 ATTGCTGGTT 155341

RESULT 16
AL512489/c
LOCUS       AL512489  174059 bp  DNA       HTG       21-MAR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-44708, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AL512489
VERSION     AL512489.9  GI:13343464
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174059)
AUTHORS     Hammond,S.
TITLE       Direct Submission
JOURNAL     Submitted (20-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 24, 2001 this sequence version replaced gi:13374819.
            -----------  Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -----------  Project Information
            Center project name: bA44708
            -----------  Summary Statistics
            Assembly program: XGAP4; version 4.5
```

```
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 173586 bases at least Q40
Consensus quality: 173783 bases at least Q30
Consensus quality: 173869 bases at least Q20
Insert size: 173959; sum-of-contigs
Insert size: 129185; 18.7% error; agarose-fp
Quality coverage: 9.22x in Q20 bases; sum-of-contigs Quality
coverage: 12.67x in Q20 bases; agarose-fp
-------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    63932: contig of 63932 bp in length
*    63933 64032: gap of    100 bp
*    64033 174059: contig of 110027 bp in length.

FEATURES          Location/Qualifiers
source            1..174059
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="6"
                  /clone_lib="RP11-44708"
                  /clone_lib="RPCI-11.2"
     misc_feature    1..63932
                  /note="assembly_fragment:00499
                  fragment_chain:1"
     misc_feature    64033..174059
                  /note="assembly_fragment:00169
                  fragment_chain:1
                  clone_end:T7
                  vector_side:right"
BASE COUNT  58218 a  30739 c  31925 g  53077 t   100 others
ORIGIN

Query Match             7.7%;   Score 34;  DB 81;  Length 174059;
Best Local Similarity   53.8%;  Pred. No. 15;
Matches  70;  Conservative  0;  Mismatches  0;  Indels  60;  Gaps  0;

Qy    107 tggaaaacagtccctgcccttttgagggtcgtctcaaacatgcccctcgagacgtcggtg 166
          ||||| || |||||||| | |||||| |||||||||| |||||||| | || |||||||
Db  81842 TGGGCAAGTTTTCAAGCCCTTCCATCTTCCACCTGGAAATGGACTCAGAGATGTTGGTG 81783

Qy    167 aagccgtagtttgtacaagactaacacgccaatggcgtcgatgttatcggagccattcg 226
          | ||| | ||| |||||| |||| |||| || ||||||  | ||| ||| ||||||
Db  81782 GGGGCACAGTGGGAATAAGACTGACTCTTTGGTTTGCATGGATGCTGGGTGAGGCCTGTG 81723

Qy    227 tgtgctgttt 236
          ||||| ||
Db  81722 ATTGCTGGTT 81713

RESULT 17
AC026746
LOCUS       AC026746  172056 bp  DNA       HTG       06-MAY-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-2017, WORKING DRAFT SEQUENCE,
            20 unordered pieces.
ACCESSION   AC026746
VERSION     AC026746.3  GI:7711977
KEYWORDS    HTG; HTGS_PHASEl; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172056)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
```

```
REFERENCE   2 (bases 1 to 172056)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On May 6, 2000 this sequence version replaced gi:7459553.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 407486
            Center clone name: RPCI-11_2017
            -----
            Summary Statistics
            Consensus quality: 155445 bases at least Q40
            Consensus quality: 165355 bases at least Q30
            Consensus quality: 166901 bases at least Q20
            Estimated insert size: 163000; pulse field gel estimation
            Estimated insert size: 170156; sum-of-contigs estimation
            Quality coverage:  3.56 in Q20 bases; pulse field gel estimation
            Quality coverage:  3.41 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 20 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1       2519: contig of 2519 bp in length
            *     2520       2619: gap of unknown length
            *     2620       5147: contig of 2528 bp in length
            *     5148       5247: gap of unknown length
            *     5248       8011: contig of 2764 bp in length
            *     8012       8111: gap of unknown length
            *     8112      13189: contig of 5078 bp in length
            *    13190      13289: gap of unknown length
            *    13290      18106: contig of 4817 bp in length
            *    18107      18206: gap of unknown length
            *    18207      22453: contig of 4247 bp in length
            *    22454      22553: gap of unknown length
            *    22554      27136: contig of 4583 bp in length
            *    27137      27236: gap of unknown length
            *    27237      31868: contig of 4632 bp in length
            *    31869      31968: gap of unknown length
            *    31969      38442: contig of 6474 bp in length
            *    38443      38542: gap of unknown length
            *    38543      45656: contig of 7114 bp in length
            *    45657      45756: gap of unknown length
            *    45757      51648: contig of 5892 bp in length
            *    51649      51748: gap of unknown length
            *    51749      58894: contig of 7146 bp in length
            *    58895      58994: gap of unknown length
            *    58995      66465: contig of 7471 bp in length
            *    66466      66565: gap of unknown length
            *    66566      73597: contig of 7032 bp in length
            *    73598      73697: gap of unknown length
            *    73698      81332: contig of 7635 bp in length
            *    81333      81432: gap of unknown length
            *    81433      90034: contig of 8602 bp in length
            *    90035      90134: gap of unknown length
            *    90135     102430: contig of 12296 bp in length
            *   102431     102530: gap of unknown length
            *   102531     116661: contig of 14131 bp in length
            *   116662     116761: gap of unknown length
            *   116762     138100: contig of 21339 bp in length
            *   138101     138200: gap of unknown length
            *   138201     172056: contig of 33856 bp in length
FEATURES    Location/Qualifiers
     source  1..172056
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="5"
             /clone="RP11-2017"
             /clone_lib="RPCI human BAC library 11"

BASE COUNT   48918 a   35715 c   36884 g   48634 t   1905 others
ORIGIN

Query Match           7.6%;  Score 33.8;  DB 70;  Length 172056;
Best Local Similarity  49.7%;  Pred. No. 17;
Matches   86;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    248 ctgtgatgacttcctgctgtactccgcttggagccattgagtgctgtttaccgtcgcagc 307
          |||-||||-|||-||||||||||-|-||||-|||-||||||||-|||||-|-||||| |||
Db 102038 CCTGCTGACACCTTAACTTTAGTCCAGTAAGACCCATTTTGGCCACCAGAAC 102097

Qy    308 actcgaagactccctgctgtactcgctggcagcatagtcgttgctgcctggtctggg 367
          ||||-||||-|-|||-||||-||-|-||||||||||||-||||-|-|||||-|-||||
Db 102098 AGTAGGAGAACACATTTGTGTTTTTTAAAGCCATTTAGAACAGTACTGGCAAGTACTGT 102157

Qy    368 gacacctgtgtactacagtagtcgttccctttgagtggtactacttctta 420
          ||-|-|||| |||-|||| ||||| |||| || ||-|-|||| ||-||-|-|
Db 102158 TACAAGTAATTAGCACAAACATTGGTTACCGTTATTATTGTTGTGTTGTTATTA 102210

RESULT 18
AL162256
LOCUS       AL162256   179556 bp   DNA             HTG        20-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-290M5, *** SEQUENCING IN
            PROGRESS ***, 7 unordered pieces.
ACCESSION   AL162256
VERSION     AL162256.5  GI:8919236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 179556)
AUTHORS     Plumb,B.
TITLE       Direct Submission
JOURNAL     Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 5, 2000 this sequence version replaced gi:8176764.
            -----
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -----     Project Information
            Center project name: bA290M5
            -----     Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 176699 bases at least Q40
            Consensus quality: 177903 bases at least Q30
            Consensus quality: 178554 bases at least Q20
            Insert size: 178956; sum-of-contigs
            Insert size: 171442; 3.5% error; agarose-fp
            Quality coverage: 4.70x in Q20 bases; sum-of-contigs Quality
            coverage: 4.91x in Q20 bases; agarose-fp
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1       5604: contig of 5604 bp in length
            *     5605       5704: gap of     100 bp
```

```
*          5705    15344: contig of 9640 bp in length
*         15345   15344: gap of     100 bp
*         15445   25851: contig of 10407 bp in length
*         25852   25951: gap of     100 bp
*         25952   77779: contig of 51828 bp in length
*         77780   77879: gap of     100 bp
*         77880   83736: contig of 5857 bp in length
*         83737   83836: gap of     100 bp
*         83837  151983: contig of 68147 bp in length
*        151984  152083: gap of     100 bp
*        152084  179556: contig of 27473 bp in length.
FEATURES             Location/Qualifiers
     source          1..179556
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone_lib="RPCI-11-290M5"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..5604
                     /note="assembly_fragment:00578
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5705..15344
                     /note="assembly_fragment:00324"
     misc_feature    15445..25851
                     /note="assembly_fragment:01062"
     misc_feature    83837..151983
                     /note="assembly_fragment:01541"
     misc_feature    152084..179556
                     /note="assembly_fragment:00360
                     clone_end:T7
                     vector_side:right"

BASE COUNT     55991 a     35593 c     36011 g     50961 t      600 others
ORIGIN

Query Match         7.6%;   Score 33.8;  DB 79;  Length 179556;
Best Local Similarity   60.2%;  Pred. No. 17;
Matches   56;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy    284 tttgaatgctgttaccgtgcagcactcgaagacttccttgctgtactcgctggcagcat 343
          | |||  ||||| ||  ||  ||| |||||  |||| | ||| | ||||| |||  |||||
Db  54035 TCTGACCGCTGCTGCCCGTAAAGGCCTCAATGAATTCCTCTCTGAGCTTCCAGACTTCCT 54094

Qy    344 agtctgtttgctgcctggtctggggacacctgt 376
          |  ||| | || | ||| || ||| | | |||||
Db  54095 TGGCTGGTGACTTGTTCCTCTTGGGACAGCTTT 54127
```

RESULT 19

```
LOCUS       AC016645   189470 bp    DNA                HTG       06-MAY-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-54A24, WORKING DRAFT SEQUENCE,
            14 unordered pieces.
ACCESSION   AC016645
VERSION     AC016645.3  GI:7711587
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189470)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189470)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint
```

```
COMMENT     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On May 6, 2000 this sequence version replaced gi:6862823.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 427125
            Center clone name: RPCI-11_54A24
            -----
            Summary Statistics
            Consensus quality: 165040 bases at least Q40
            Consensus quality: 177146 bases at least Q30
            Consensus quality: 181696 bases at least Q20
            Estimated insert size: 182000; pulse field gel estimation
            Estimated insert size: 188170; sum-of-contigs estimation
            Quality coverage:  4.74 in Q20 bases; pulse field gel estimation
            Quality coverage:  4.59 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    3131: contig of 3131 bp in length
            *     3132    3231: gap of unknown length
            *     3232    6853: contig of 3622 bp in length
            *     6854    6953: gap of unknown length
            *     6954   10621: contig of 3668 bp in length
            *    10622   10721: gap of unknown length
            *    10722   15235: contig of 4514 bp in length
            *    15236   15335: gap of unknown length
            *    15336   21857: contig of 6522 bp in length
            *    21858   21957: gap of unknown length
            *    21958   32538: contig of 10581 bp in length
            *    32539   32638: gap of unknown length
            *    32639   41798: contig of 9160 bp in length
            *    41799   41898: gap of unknown length
            *    41899   53135: contig of 11237 bp in length
            *    53136   53235: gap of unknown length
            *    53236   64815: contig of 11580 bp in length
            *    64816   64915: gap of unknown length
            *    64916   77781: contig of 12866 bp in length
            *    77782   77881: gap of unknown length
            *    77882   92170: contig of 14289 bp in length
            *    92171   92270: gap of unknown length
            *    92271  111202: contig of 18932 bp in length
            *   111203  111302: gap of unknown length
            *   111303  143824: contig of 32522 bp in length
            *   143825  143924: gap of unknown length
            *   143925  189470: contig of 45546 bp in length.
FEATURES             Location/Qualifiers
     source          1..189470
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-54A24"
                     /clone_lib="RPCI human BAC library 11"

BASE COUNT     54627 a     38681 c     41992 g     52840 t     1330 others
ORIGIN

Query Match         7.6%;   Score 33.8;  DB 64;  Length 189470;
Best Local Similarity   49.7%;  Pred. No. 17;
Matches   86;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy    248 cttgatgacttccttgctgtactccgcttggagccatttgagtgctgtttaccgtgcagc 307
          | ||||| ||||| || || ||  || |  || || |||| | ||||||  |||| |||
Db  34990 CCTGCTGACACCTTAACTTAGTCCAGTAAGACCCATTTGGAGTTTTGGCCACCAGAAC 35049
```

```
Qy    308  actcgaagacttccttgctgtactcgctggcagcatagtctgtttgctgcctggtctggg 367
            ||  |||||| || |||  ||       ||   ||  ||| || ||| | ||||
Db  35050  AGTAGGAGAACACATTTGTGTTTTTAAAGCCATTAGAACAGTACTTGGCAAGTACTGT 35109

Qy    368  gacacctgtactacagtactcgttccctttgtagtggtactaacttctta 420
            ||| ||  || ||||   || |||  ||    || |||||||| | |||
Db  35110  TACAAGTAATTAGCACAAACATTGGTTTACCGTTATTATTGTTGTTGTTATTA 35162

RESULT 20
AC073540/c
LOCUS       AC073540    189942 bp    DNA                  HTG        31-AUG-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-164A5, WORKING DRAFT SEQUENCE,
            16 unordered pieces.
ACCESSION   AC073540
VERSION     AC073540.2  GI:9954696
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189942)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 189942)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            -----Genome Center
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 469346
            Center clone name: RPCI-11_164A5
            -----
COMMENT     Summary Statistics
            Consensus quality: 176984 bases at least Q40
            Consensus quality: 184449 bases at least Q30
            Consensus quality: 185674 bases at least Q20
            Estimated insert size: 18450; agarose-fp estimation
            Estimated insert size: 18442; sum-of-contigs estimation
            Quality coverage:  6.05 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.93 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1  1317: contig of 1317 bp in length
            *    1318  1417: gap of unknown length
            *    1418  2526: contig of 1109 bp in length
            *    2527  2626: gap of unknown length
            *    2627  5034: contig of 2408 bp in length
            *    5035  5134: gap of unknown length
            *    5135  7507: contig of 2373 bp in length
            *    7508  7607: gap of unknown length
            *    7608 10242: contig of 2635 bp in length
            *   10243 10342: gap of unknown length
            *   10343 14775: contig of 4433 bp in length
            *   14776 14875: gap of unknown length
            *   14876 20332: contig of 5457 bp in length
            *   20333 20432: gap of unknown length
            *   20433 27996: contig of 7564 bp in length
            *   27997 28096: gap of unknown length
            *   28097 39155: contig of 11059 bp in length
```

```
            *   39156 39255: gap of unknown length
            *   39256 56163: contig of 16908 bp in length
            *   56164 56263: gap of unknown length
            *   56264 72833: contig of 16570 bp in length
            *   72834 72933: gap of unknown length
            *   72934 88357: contig of 15424 bp in length
            *   88358 88457: gap of unknown length
            *   88458 103315: contig of 14858 bp in length
            *  103316 103415: gap of unknown length
            *  103416 127722: contig of 24307 bp in length
            *  127723 127822: gap of unknown length
            *  127823 155816: contig of 27994 bp in length
            *  155817 155916: gap of unknown length
            *  155917 189942: contig of 34026 bp in length.
FEATURES             Location/Qualifiers
     source          1..189942
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-164A5"
                     /clone_lib="RPCI human BAC library 11"
BASE COUNT    54100 a   40736 c   39831 g   53756 t   1519 others
ORIGIN

Query Match          7.6%;  Score 33.8;  DB 74;  Length 189942;
Best Local Similarity  49.7%;  Pred. No. 17;
Matches   86;  Conservative   0;  Mismatches   87;  Indels   0;  Gaps   0;

Qy    248  cttgatgacttccttgctgtactccgcttggagccattgagtgctgttaccgtgcagc 307
            || ||||| |||| ||| |||| |||  ||| |||   || ||||  ||| |||| |||
Db  172334  CCTGCTGACACCTTAACTTAGTCCAGTAAGACCCATTTGGAGTTTGGCCACCAGAAC 172275

Qy    308  actcgaagacttccttgctgtactcgctggcagcatagtctgtttgctgcctggtctggg 367
            || |||| || ||||  ||   ||  |||  ||  ||| || ||| || ||| | |||
Db  172274  AGTAGGAGAACACATTTGTGTTTTTTAAAGCCATTAGAACAGTACTTGGCAAGTACTGT 172215

Qy    368  gacacctgtactacagtactcgttccctttgtagtggtactaacttctta 420
            ||| ||  || ||||   || |||  ||    || |||||||| | |||
Db  172214  TACAAGTAATTAGCACAAACATTGGTTTACCGTTATTATTGTTGTTGTTATTA 172162

RESULT 21
AC026750/c
LOCUS       AC026750    196716 bp    DNA                  HTG        26-SEP-2000
DEFINITION  Homo sapiens chromosome 5 clone RP11-469O13, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AC026750
VERSION     AC026750.3  GI:10305172
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196716)
AUTHORS     DOE Joint Genome Institute.
TITLE       Sequencing of Human Chromosome 5
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 196716)
AUTHORS     DOE Joint Genome Institute.
TITLE       Direct Submission
JOURNAL     Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            On Sep 26, 2000 this sequence version replaced gi:7459583.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 586810
            Center clone name: RPCI-11_469O13
            -----
```

```
Db 179108  |||||||| |||| ||| |||| |||||| |||||||||||||||||| |||| |||||
           CCTGCTGACACCTTAACTTTAGTCCAGTAAGACCCATTTTGGCACCAGAAC  179049
Qy    308  actcgaagactttcctgctgactcgtcggcgcagatcgtcgtttgctgcctggtctggg  367
           |||| ||| ||| ||| |||| ||  ||| || | ||| || ||| ||
Db 179048  AGTAGGAGAACACATTTGTGTTTTTTTAAAGCCATTGAACAGTACTTGGCAAGTACTGT  178989
Qy    368  gacacctgtactacagtactcgttccctttgtagtggtactacttccta  420
           |||| ||| ||||| |||| |||| ||||||| |||
Db 178988  TACAAGTAATTAGCACAAACATTGGTTTACCCTTATTATTGTGTTGTTATTA  178936
```

RESULT 22
ACO26809
LOCUS   ACO26809   213351 bp   DNA   HTG   26-SEP-2000
DEFINITION   Homo sapiens chromosome 5 clone RP11-54J1, WORKING DRAFT SEQUENCE,
           21 unordered pieces.
ACCESSION   ACO26809
VERSION   ACO26809.2   GI:10305178
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
   ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 213351)
   AUTHORS   DOE Joint Genome Institute.
   TITLE   Sequencing of Human Chromosome 5
   JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 213351)
   AUTHORS   DOE Joint Genome Institute.
   TITLE   Direct Submission
   JOURNAL   Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   On Sep 26, 2000 this sequence version replaced gi:7321454.
           ------Genome Center
           Center: Joint Genome Institute
           Center Code: JGI
           Web site: http://www.jgi.doe.gov

           Project Information
           Center Project Name: 427318
           Center clone name: RPCI-11_54J1

           Summary Statistics
           Consensus quality: 199956 bases at least Q40
           Consensus quality: 206278 bases at least Q30
           Consensus quality: 208240 bases at least Q20
           Estimated insert size: 361620; agarose-fp estimation
           Estimated insert size: 211351; sum-of-contigs estimation
           Quality coverage: 8.7 in Q20 bases; agarose-fp estimation
           Quality coverage: 14.89 in Q20 bases; sum-of-contigs estimation
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 21 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *   1   1195: contig of 1195 bp in length
          *   1196   1295: gap of unknown length
          *   1296   2320: contig of 1025 bp in length
          *   2321   2420: gap of unknown length
          *   2421   3651: contig of 1231 bp in length
          *   3652   3751: gap of unknown length
          *   3752   5013: contig of 1262 bp in length
          *   5014   5113: gap of unknown length
          *   5114   6225: contig of 1112 bp in length
          *   6226   6325: gap of unknown length
          *   6326   7394: contig of 1069 bp in length
          *   7395   7494: gap of unknown length
          *   7495   8516: contig of 1022 bp in length
          *   8517   8616: gap of unknown length

Summary Statistics
Consensus quality: 181431 bases at least Q40
Consensus quality: 190016 bases at least Q30
Consensus quality: 191521 bases at least Q20
Estimated insert size: 184750; agarose-fp estimation
Estimated insert size: 194716; sum-of-contigs estimation
Quality coverage: 3.35 in Q20 bases; agarose-fp estimation
Quality coverage: 3.18 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*   1   1027: contig of 1027 bp in length
*   1028   1127: gap of unknown length
*   1128   2214: contig of 1087 bp in length
*   2215   2314: gap of unknown length
*   2315   3438: contig of 1124 bp in length
*   3439   3538: gap of unknown length
*   3539   5097: contig of 1559 bp in length
*   5098   5197: gap of unknown length
*   5198   6674: contig of 1477 bp in length
*   6675   6774: gap of unknown length
*   6775   8091: contig of 1317 bp in length
*   8092   8191: gap of unknown length
*   8192   9300: contig of 1109 bp in length
*   9301   9400: gap of unknown length
*   9401   11808: contig of 2408 bp in length
*   11809   11908: gap of unknown length
*   11909   14543: contig of 2635 bp in length
*   14544   14643: gap of unknown length
*   14644   17016: contig of 2373 bp in length
*   17017   17116: gap of unknown length
*   17117   21549: contig of 4433 bp in length
*   21550   21649: gap of unknown length
*   21650   27106: contig of 5457 bp in length
*   27107   27206: gap of unknown length
*   27207   34770: contig of 7564 bp in length
*   34771   34870: gap of unknown length
*   34871   45929: contig of 11059 bp in length
*   45930   46029: gap of unknown length
*   46030   62937: contig of 16908 bp in length
*   62938   63037: gap of unknown length
*   63038   79607: contig of 16570 bp in length
*   79608   79707: gap of unknown length
*   79708   95131: contig of 15424 bp in length
*   95132   95231: gap of unknown length
*   95232   110089: contig of 14858 bp in length
*   110090   110189: gap of unknown length
*   110190   134496: contig of 24307 bp in length
*   134497   134596: gap of unknown length
*   134597   162590: contig of 27994 bp in length
*   162591   162690: gap of unknown length
*   162691   196716: contig of 34026 bp in length.
FEATURES   Location/Qualifiers
   source   1..196716
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-469013"
            /clone_lib="RPCI human BAC library 11"
BASE COUNT   54117 a   40760 c   39860 g   53767 t   8212 others
ORIGIN

Query Match   7.6%;   Score 33.8;   DB 70;   Length 196716;
Best Local Similarity   49.7%;
Matches   86;   Conservative   0;   Mismatches   87;   Indels   0;   Gaps   0;

Qy   248   cttgatgacttcctttgctgctactccgcttggagccattggagtgctgttaccgtgcagc   307

```
*          8617    9697: contig of 1081 bp in length
*          9698    9797: gap of unknown length
*         9798   11178: contig of 1381 bp in length
*        11179   11278: gap of unknown length
*        11279   12280: contig of 1002 bp in length
*        12281   12380: gap of unknown length
*        12381   13779: contig of 1399 bp in length
*        13780   13879: gap of unknown length
*        13880   14885: contig of 1006 bp in length
*        14886   14985: gap of unknown length
*        14986   16581: contig of 1596 bp in length
*        16582   16681: gap of unknown length
*        16682   17804: contig of 1123 bp in length
*        17805   17904: gap of unknown length
*        17905   19230: contig of 1325 bp in length
*        19230   19329: gap of unknown length
*        19330   21164: contig of 1835 bp in length
*        21165   21264: gap of unknown length
*        21265   22326: contig of 1062 bp in length
*        22327   22426: gap of unknown length
*        22427   23714: contig of 1288 bp in length
*        23715   23814: gap of unknown length
*        23815   25625: contig of 1811 bp in length
*        25626   25725: gap of unknown length
*        25726   27489: contig of 1764 bp in length
*        27490   27589: gap of unknown length
*        27590  213351: contig of 185762 bp in length.

FEATURES          Location/Qualifiers
source          1..213351
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="5"
                /clone_lib="RPCI human BAC library 11"
                /clone="RP11-54J1"

BASE COUNT    53780 a  39729 c  41515 g  55288 t  23039 others
ORIGIN

Query Match           7.6%;    Score 33.8;  DB 70;  Length 213351;
Best Local Similarity  49.7%;  Pred. No. 17;
Matches  86;  Conservative  0;  Mismatches  87;  Indels  0;  Gaps  0;

Qy    248  cttgatgacttcctgctgtactccgcttggagccattggagtgctgttaccgtgcagc  307
           |||||||| ||| |||||| || | ||| |||  | |||||| | |  ||||   || |
Db 184277  CCTGCTGCACACCTTAACTTTAGTCCAGTAAGACCCATTTTGGAGTTTTGGCCACCAGAAC 184336

Qy    308  actcgaagacttcctgctgctactcgctggcagcatagtcgtttgctgcctggtctggg  367
           ||  |   |||  | |||  | |  ||||||| | | ||||   | ||| | |||| |||
Db 184337  AGTAGGAGAAACACATTTGTGTTTTTAAAGCCATTTAGAACAGTACTTGGCAAGTACTGT  184396

Qy    368  gacacctgtactacagtactctcgttccctttgtagtggtactacttcctta  420
           |||  || ||| |||  ||| ||  || ||||||||||  ||||  |  ||||
Db 184397  TACAAGTAATTAGCACAAACATTGGTTTACCGTTATTATTATTGTTGTTGTATTA  184449


RESULT  23
HUMF13A14
LOCUS       HUMF13A14     273 bp     DNA             PRI       08-NOV-1994
DEFINITION  Human coagulation factor XIII a subunit gene, exon 14.
ACCESSION   M22000 J03834
VERSION     M22000.1  GI:182306
KEYWORDS    factor XIII; fibrinoligase; glycoprotein; plasma transglutaminase.
SEGMENT     14 of 15
SOURCE      Human DNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 273)
  AUTHORS   Ichinose,A. and Davie,E.W.
  TITLE     Characterization of the gene for the a subunit of human factor XIII
            (plasma transglutaminase), a blood coagulation factor
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (16), 5829-5833 (1988)
  MEDLINE   88320337
```

```
FEATURES          Location/Qualifiers
source          1..273
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="6p25-p24"
     intron     <1..52
                /gene="F13A1"
                /note="F13A1, intron M"
     prim_transcript  <1..>273
                /gene="F13A1"
                /note="F13A1 mRNA and introns"
     exon       53..189
                /gene="F13A1"
                /note="G00-120-614"
                /number=14
     intron     190..>273
                /gene="F13A1"
                /note="F13A1, intron N"

BASE COUNT    67 a  56 c  74 g  76 t
ORIGIN        Unknown number of bp after segment 13.

Query Match           7.6%;    Score 33.6;  DB 97;  Length 273;
Best Local Similarity  59.4%;  Pred. No. 18;
Matches  57;  Conservative  0;  Mismatches  39;  Indels  0;  Gaps  0;

Qy    135  tcgtctcaaacatgccctgcagacgtcggtgaagccgtagttgtacaagactacaac  194
           |||| ||| ||||  || | || | |||| |||| ||| |||  |  ||||| |||||
Db    109  TCCTTTAAAGAAACCCTGCGAAATGTCTGGGTACACCTGATGGTCCTGGAGTAACAAG  168

Qy    195  gccaatggcgtcgatgttatcggagccattcgtgtg  230
           ||||| | |||||  | ||| |||| | ||| || ||
Db    169  ACCAATGAAGAAGATGTTCCGTAAACCTTGGGTTG  204


RESULT  24
AC023009
LOCUS       AC023009    150349 bp    DNA            HTG       03-MAR-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-1002 map 6, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC023009
VERSION     AC023009.3  GI:7145007
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150349)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-1002
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150349)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,K.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassillev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 3, 2000 this sequence version replaced gi:7139816.
```

```
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----         Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
----         Project Information
             Center project name: L6612
             Center clone name: 10_O_2
----         Summary Statistics
             Sequencing vector: M13; M77815; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.960731
             Consensus quality: 140362 bases at least Q40
             Consensus quality: 145559 bases at least Q30
             Consensus quality: 147640 bases at least Q20
             Insert size: 145000; agarose-fp
             Insert size: 149349; sum-of-contigs
             Quality coverage: 4.3 in Q20 bases; agarose-fp
             Quality coverage: 4.2 in Q20 bases; sum-of-contigs
----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1..  1051: contig of 1051 bp in length
* 1052   1151: gap of    100 bp
* 1152    3060: contig of 1909 bp in length
* 3061   3160: gap of    100 bp
* 3161    6163: contig of 3003 bp in length
* 6164   6263: gap of    100 bp
* 6264    8624: contig of 2361 bp in length
* 8625   8724: gap of    100 bp
* 8725   16968: contig of 8244 bp in length
* 16969  17068: gap of    100 bp
* 17069  27598: contig of 10530 bp in length
* 27599  27698: gap of    100 bp
* 27699  38339: contig of 10641 bp in length
* 38340  38439: gap of    100 bp
* 38440  59332: contig of 20893 bp in length
* 59333  59432: gap of    100 bp
* 59433  79663: contig of 20231 bp in length
* 79664  79763: gap of    100 bp
* 79764  109782: contig of 30019 bp in length
* 109783 109882: gap of    100 bp
* 109883 150349: contig of 40467 bp in length.
FEATURES             Location/Qualifiers
     source          1..150349
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-1002"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1051
                     /note="assembly_fragment"
     misc_feature    1152..3060
                     /note="assembly_fragment"
     misc_feature    3161..6163
                     /note="assembly_fragment"
     misc_feature    6264..8624
                     /note="assembly_fragment"
     misc_feature    8725..16968
                     /note="assembly_fragment"
     misc_feature    17069..27598
                     /note="assembly_fragment"
     misc_feature    27699..38339
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    38440..59332
                     /note="assembly_fragment"
     misc_feature    59433..79663
                     /note="assembly_fragment"
     misc_feature    79764..109782
                     /note="assembly_fragment"
     misc_feature    109883..150349
                     /note="assembly_fragment"
BASE COUNT    43365 a   30002 c   31050 g   44930 t     1002 others
ORIGIN

Query Match                7.6%;   Score 33.6;   DB 67;   Length 150349;
Best Local Similarity     59.4%;   Pred. No. 20;
Matches  57;  Conservative  0;  Mismatches  39;  Indels  0;  Gaps  0;

Qy   135  tcgtctcaaacatgcccctcgagacgtcggtgaagccgtagtttgtacaagactaaacaac 194
          ||||||||||||||||-|||-|||||||||||||||||||-|||||||| -| |||||
Db 118258  TCCTTAAAGAAGAAACCCTGCGAAATGTCTGGGTACACCTGGGATGGTCCTGGAGTAACAAG 118317

Qy   195  gccaatggcgtcgatgttatcggagccattcgtgtg 230
          |||||| |  |||||| |||||||  |||| | |||||
Db 118318  ACCAATGAAGAAGATGTTCCGGTAAACCTTGGGTTG 118353

RESULT 25
AL157775/c
LOCUS       AL157775    163960 bp    DNA                 PRI      12-APR-2001
DEFINITION  Human DNA sequence from clone RP11-525O21 on chromosome 6 Contains
            the 3' end of the F13A1 (coagulation factor XIII, A1 polypeptide)
            gene, ESTs, STSs and GSSs, complete sequence.
ACCESSION   AL157775
VERSION     AL157775.15  GI:11121004
KEYWORDS    HTG; coagulation factor; F13A1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 163960)
AUTHORS     Sehra,H.
TITLE       Direct Submission
JOURNAL     Submitted (29-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:10932189.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-525O21 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The right end of clone RP11-525O21 is at 163960 in this
            sequence. The true left end of clone RP11-232H4 is at 132449 in
            this sequence. The true right end of clone RP3-380B8 is at 100 in
            this sequence. This sequence was finished as follows unless
            otherwise noted: all regions were either double-stranded or
            sequenced with an alternate chemistry or covered by high quality
            data (i.e., phred quality >= 30); an attempt was made to resolve
```

```
                all sequencing problems, such as compressions and repeats; all
                regions were covered by at least one plasmid subclone or more than
                one M13 subclone; and the assembly was confirmed by restriction
                digest. RP11-525O21 is from the library RPCI-11.2 constructed by
                the group of Pieter de Jong. For further details see
                http://www.chori.org/bacpac/home.htm
                VECTOR: pACe3.6.
FEATURES        Location/Qualifiers
     source     1..163960
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /clone="RP11-525O21"
                /clone_lib="RPCI-11.2"
     repeat_region   3535..3669
     misc_feature    complement(3583..3911)
                /note="3 copies 45 mer 85% conserved"
     misc_feature    3630..3907
                /note="match: GSS: Em:AQ705901"
     misc_feature    3664..3777
                /note="match: GSS: Em:AQ253322"
     misc_feature    3676..3843
                /note="match: GSS: Em:AQ317328"
     misc_feature    complement(3684..3839)
                /note="match: GSS: Em:B50322"
     misc_feature    3685..3926
                /note="match: GSS: Em:AQ530998"
     repeat_region   3730..3759
                /note="match: GSS: Em:AQ526386"
     repeat_region   3732..3759
                /note="15 copies 2 mer at 96% conserved"
     repeat_region   3812..3839
                /note="7 copies 4 mer atat 100% conserved"
     repeat_region   3814..3837
                /note="14 copies 2 mer at 96% conserved"
     repeat_region   4094..4408
                /note="6 copies 4 mer atat 100% conserved"
     misc_feature    complement(4428..4749)
                /note="7 copies 45 mer 74% conserved"
     repeat_region   4491..4580
                /note="match: STS: Em:HS451L15S"
     repeat_region   4551..4614
                /note="2 copies 45 mer 87% conserved"
     misc_feature    4565..5092
                /note="32 copies 2 mer at 71% conserved"
     repeat_region   4730..5241
                /note="match: GSS: Em:AQ371839"
     repeat_region   5435..5547
                /note="L2 repeat: matches 1749. .2282 of consensus"
     repeat_region   5601..5680
                /note="L2 repeat: matches 2590. .2749 of consensus"
     repeat_region   6076..6362
                /note="L2 repeat: matches 1299. .1380 of consensus"
     repeat_region   6802..7103
                /note="AluY repeat: matches 1. .285 of consensus"
     repeat_region   7329..7446
                /note="L1 repeat: matches 4734. .5047 of consensus"
     repeat_region   7520..8075
                /note="59 copies 2 mer tt 66% conserved"
     misc_feature    11177..11578
                /note="L1MC5 repeat: matches 7354. .7935 of consensus"
     repeat_region   11585..11651
                /note="match: GSS: Em:B71797"
     misc_feature    11606..11887
                /note="MER5A repeat: matches 9. .75 of consensus"
     repeat_region   11655..11738
                /note="match: STS: Em:G17084"
     repeat_region   12974..13172
                /note="AluS repeat: matches 1. .84 of consensus"
     repeat_region   13344..13419
                /note="L2 repeat: matches 2214. .2419 of consensus"
     misc_feature    13560..13941
                /note="19 copies 4 mer aaga 65% conserved"

     repeat_region   13994..14175
                /note="match: GSS: Em:AQ005912"
     repeat_region   17440..17721
                /note="MER5A repeat: matches 3. .178 of consensus"
     repeat_region   17831..18376
                /note="L1MC5 repeat: matches 7628. .7920 of consensus"
     repeat_region   18377..18677
                /note="L1PB1 repeat: matches 5599. .6155 of consensus"
     repeat_region   18678..19023
                /note="AluSc repeat: matches 1. .295 of consensus"
     repeat_region   19191..19240
                /note="L1PB1 repeat: matches 5257. .5599 of consensus"
     repeat_region   19241..19540
                /note="L1M4c repeat: matches 1685. .1737 of consensus"
     repeat_region   19541..19686
                /note="AluY repeat: matches 1. .300 of consensus"
     repeat_region   21482..21897
                /note="L1M4c repeat: matches 1555. .1685 of consensus"
     repeat_region   22715..22790
                /note="L1PA10 repeat: matches 5743. .6161 of consensus"
     repeat_region   22791..25265
                /note="L1PA6 repeat: matches 6066. .6143 of consensus"
     repeat_region   25267..25933
                /note="HERVK_9 repeat: matches 1625. .4106 of consensus"
     repeat_region   25934..26423
                /note="HERVK_9 repeat: matches 1. .661 of consensus"
     repeat_region   26460..26775
                /note="MER9 repeat: matches 1. .511 of consensus"
     repeat_region   26776..27304
                /note="AluYa5 repeat: matches 1. .311 of consensus"
     repeat_region   27305..27615
                /note="L1P repeat: matches 3090. .3619 of consensus"
     repeat_region   27616..30704
                /note="AluY repeat: matches 1. .309 of consensus"
     repeat_region   30924..31188
                /note="L1P repeat: matches 11. .3090 of consensus"
     repeat_region   31832..32363
                /note="AluY repeat: matches 9. .291 of consensus"
     repeat_region   32365..32535
                /note="MER74A repeat: matches 20. .558 of consensus"
     repeat_region   32632..32776
                /note="L1MA4 repeat: matches 5942. .6118 of consensus"
     repeat_region   32778..33125
                /note="MER47A repeat: matches 5. .150 of consensus"
     repeat_region   33187..33689
                /note="L1MEC repeat: matches 2404. .2399 of consensus"
     repeat_region   34054..34446
                /note="L1M4 repeat: matches 2645. .3148 of consensus"
     repeat_region   34511..37154
                /note="L1 repeat: matches 3547. .3949 of consensus"
     repeat_region   37353..37603
                /note="L1MA2 repeat: matches 3652. .6308 of consensus"
     repeat_region   38056..38250
                /note="L1M4 repeat: matches 4795. .5055 of consensus"
     repeat_region   38350..38377
                /note="MIR repeat: matches 50. .227 of consensus"
     repeat_region   38352..38375
                /note="14 copies 2 mer ac 96% conserved"
     repeat_region   38759..39104
                /note="6 copies 4 mer acac 100% conserved"
     repeat_region   39449..39487
                /note="L2 repeat: matches 1770. .2112 of consensus"
     misc_feature    40297..40706
                /note="match: GSS: Em:AQ176196"
     repeat_region   40494..40895
                /note="MLT1J repeat: matches 111. .149 of consensus"
     repeat_region   41695..42062
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   43042..43469
                /note="MLT1A2 repeat: matches 1. .370 of consensus"
     repeat_region   43605..43763
                /note="L1PB2 repeat: matches 5711. .6155 of consensus"
     repeat_region   43605..43763
                /note="MER5A repeat: matches 12. .189 of consensus"
```

```
repeat_region    44885..44994
                 /note="L2 repeat: matches 2593..2690 of consensus"
repeat_region    44997..45420
                 /note="MLT2FB repeat: matches 1..414 of consensus"
repeat_region    46331..46356
                 /note="13 copies 2 mer tg 92% conserved"
repeat_region    46597..46636
                 /note="20 copies 2 mer ca 82% conserved"
repeat_region    46612..46635
                 /note="6 copies 4 mer acac 100% conserved"
repeat_region    46711..46754

Query Match              7.6%;     Score 33.6; DB 90;  Length 163960;
Best Local Similarity   59.4%;     Pred. No. 20;
Matches  57;  Conservative   0;  Mismatches  39;  Indels  0;  Gaps  0;

Qy    135 tcgtcccaaacatgccccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
          |||-|| |||| ||||||||| | || |||||||||||||||||| | ||||||
Db 109722 TCCTTAAAAGAAACCCTGCGAAATGTCTGGGTACACCTCGATGGTCCTCGAGTAACAAG 109663

Qy    195 gccaatgggcgtgatgttatcggagccattcgtgtg 230
          |||||| |||||||| ||||||| ||| || |||
Db 109662 ACCAATGAAGAAGATGTTCCGGTAAACCTTGGGTTG 109627

RESULT 26
AC025972
LOCUS       AC025972    166402 bp    DNA          HTG     12-JAN-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-6K5 map 11, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC025972
VERSION     AC025972.4  GI:12084073
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166402)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 11 clone RP11-6K5
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 166402)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Illev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (18-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 12, 2001 this sequence version replaced gi:8076808.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------- project information
            Center project name: L6107
            Center clone name: 6_K_5
            ---------- Summary Statistics
            Sequencing vector: M13; M77815; 40% of reads
            Sequencing vector: Plasmid: n/a; 60% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 165502 bases at least Q40
            Consensus quality: 165912 bases at least Q30
            Consensus quality: 166038 bases at least Q20
            Insert size: 172000; agarose-fp
            Insert size: 166202; sum-of-contigs
            Quality coverage: 10.5 in Q20 bases; agarose-fp
            Quality coverage: 10.9 in Q20 ba.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1 137540: contig of 137540 bp in length
          * 137541 137640: gap of 100 bp
          * 137641 137742: contig of 102 bp in length
          * 137743 137842: gap of 100 bp
          * 137843 166402: contig of 28560 bp in length.
FEATURES             Location/Qualifiers
     source          1..166402
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-6K5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..137540
                     /note="assembly_fragment"
                     clone_end:SP6
                     vector_side:left"
     misc_feature    137641..137742
                     /note="assembly_fragment"
     misc_feature    137843..166402
                     /note="assembly_fragment"
                     clone_end:T7
                     vector_side:right"
BASE COUNT   49603 a   32681 c   33231 g   50687 t    200 others
ORIGIN

Query Match              7.6%;     Score 33.6; DB 69;  Length 166402;
Best Local Similarity   59.4%;     Pred. No. 20;
Matches  57;  Conservative   0;  Mismatches  39;  Indels  0;  Gaps  0;

Qy    34 tagtggtcatctgaaggaaggaaggacaaagacacctccaagccgacaggctactgtagc 93
         |||||||||||||||| ||||||| | |||||| |||||||  | |||||||||||||||
Db 77827 TAGTGGGGAAACTGAAGTATAGGCATATTACAACAACAACAATGACAATGGTATGGTAGA 77886

Qy    94 tggcacggcataatgaaaacagtccctgccctttt 129
         || || |||||| ||||| |||| |||| |||
Db 77887 TTGAGTCCCATAGTGGACACCGAGTCTTCAGCCTTGT 77922

RESULT 27
AC068860
LOCUS       AC068860    188642 bp    DNA          HTG     16-MAR-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-371I20 map 11, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC068860
VERSION     AC068860.3  GI:13357505
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
```

```
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188642)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens chromosome 11, clone RP11-371I20
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 188642)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (10-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:8516115.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------ Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ------------------ Project Information
              Center project name: L8794
              Center clone name: 371_I_20
            ------------------ Summary Statistics
              Sequencing vector: M13; M77815; 72% of reads
              Sequencing vector: Plasmid; n/a; 28% of reads
              Chemistry: Dye-terminator Big Dye; 100% of reads
              Assembly program: Phrap; version 0.960731
              Consensus quality: 184202 bases at least Q40
              Consensus quality: 186774 bases at least Q30
              Consensus quality: 187632 bases at least Q20
              Insert size: 187000; agarose-fp
              Insert size: 188142; sum-of-contigs
              Quality coverage: 5.7 in Q20 bases; agarose-fp
              Quality coverage: 5.7 in Q20 b.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1  22203: contig of 22203 bp in length
            * 22204 22303: gap of    100 bp in length
            * 22304  26021: contig of 3718 bp in length
            * 26022 26121: gap of    100 bp
            * 26122  90812: contig of 64691 bp in length
            * 90813 90912: gap of    100 bp
            * 90913 125638: contig of 34726 bp in length
            * 125639 125738: gap of    100 bp
```

```
            * 125739 174788: contig of 49050 bp in length
            * 174789 174888: gap of     100 bp
            * 174889 188642: contig of 13754 bp in length.
FEATURES             Location/Qualifiers
     source          1..188642
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-371I20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..22203
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22304..26021
                     /note="assembly_fragment"
     misc_feature    26122..90812
                     /note="assembly_fragment"
     misc_feature    90913..125638
                     /note="assembly_fragment"
     misc_feature    125739..174788
                     /note="assembly_fragment"
     misc_feature    174889..188642
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    55262 a   38577 c   38103 g   56187 t    513 others
ORIGIN

Query Match            7.6%;  Score 33.6;  DB 73;  Length 188642;
Best Local Similarity 59.4%;  Pred. No. 20;
Matches  57;  Conservative  0;  Mismatches  39;  Indels  0;  Gaps  0;

Qy     34 tagtggtcatctgaaggaaggacaaggacacctccaagccggacacggctactgtagc 93
          ||||||| ||||||| |||| |||-||||-|| ||||| ||| ||||| |||| ||||
Db 182343 TAGTGGGAAACTGAAGTATAGGCATATTACAACAACAACAATGCAATGGTATGGTAGA 182402

Qy     94 tggcacggcataatggaaaacagtccctgccctt 129
          |||| |||| |||| |||||| |||  |||| |
Db 182403 TTGAGTCCCATAGTGGACACGAGTCTTCAGCCTTGT 182438


RESULT 28
AL139335
LOCUS       AL139335        281423 bp    DNA             HTG        23-JAN-2001
DEFINITION  Homo sapiens chromosome 6 clone RP3-377I1, *** SEQUENCING IN
            PROGRESS ***, 9 unordered pieces.
ACCESSION   AL139335
VERSION     AL139335.18  GI:12539502
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 281423)
AUTHORS     Skuce,C.
TITLE       Direct Submission
JOURNAL     Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 25, 2001 this sequence version replaced gi:10432514.
            ----------- Genome Center
              Center: Sanger Centre
              Center code: SC
              Web site: http://www.sanger.ac.uk
              Contact: humquery@sanger.ac.uk
            ----------- Project Information
              Center project name: dJ377I1
            ----------- Summary Statistics
              Assembly program: XGAP4; version 4.5
              Sequencing vector: M13; M77815; 1% of reads
```

```
Db 276240 TCCTTTAAAAGAAACCTGCGAAATGTCTGGGTACACCTGGATGGTCCTGGAGTAACAAG 276299
                  ||||||  | |||||| | |||| |  |||| |||
Qy    195 gccaatggcgtcgatgttatcggagccattcgtgtg 230
                  ||||||  ||||||  ||| || ||  |||| ||
Db 276300 ACCAATGAAGAGAGATGTTCCGGTAAACCTTGGGTTG 276335

RESULT 29
AX048148/c
LOCUS        AX048148    1894 bp    DNA              PAT    15-DEC-2000
DEFINITION   Sequence 1 from Patent WO0070092.
ACCESSION    AX048148
VERSION      AX048148.1    GI:11876967
KEYWORDS     .
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1894)
AUTHORS      Kaufmann,J. and Xin,H.
TITLE        Expression of ets-domain proteins in cancer
JOURNAL      Patent: WO 0070092-A 1 23-NOV-2000;
             Chiron Corporation (US)
FEATURES     Location/Qualifiers
source       1..1894
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
CDS          416..1423
             /note="Human JKETS."
             /codon_start=1
             /protein_id="CAC19002.1"
             /db_xref="GI:11876968"
             /translation="MSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
             PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREPPEEPEQCPVIDSQAPA
             GSLDLVPGGLTEEHSLEQVOSMVVGEVLKDIETACKLLNTADPMDWSPSSNVQKWLL
             WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRORSPLGGDVLHAHLDIWKSAAWMKER
             TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
             GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYKKGIIRKPDISQRLVYQFVHP
             I"

BASE COUNT     368 a     653 c     571 g     302 t
ORIGIN

Query Match      7.5%;
Best Local Similarity    58.6%;    Score 33.4;    DB 9;    Length 1894;
Matches   58;    Conservative    0;    Mismatches   41;    Indels    0;    Gaps    0;

Qy   290 tgctgttaccgtgcagcactcgaagacttccttgctgctactcgctggcagcatagtctg 349
             ||||          |  | ||||||| |||   |||  |  || ||   ||  |||
Db   443 TGCTCAGACCGGCGCTGGCGCGTGCCCATGCCGCTGCGTGTTGGGCTGGCGCTGCGTGTCTG 384

Qy   350 tttgctgcctgtctgggacacctgtgtactacagtac 388
             || |||||||| ||    || |  ||| || ||
Db   383 TTAGCTGCCTGGTGCCCAGGAGCTGTCTGCTGCAGTGC 345

RESULT 30
AF071538/c
LOCUS        AF071538    1894 bp    mRNA             PRI    10-JAN-2000
DEFINITION   Homo sapiens Ets transcription factor PDEF (PDEF) mRNA, complete
             cds.
ACCESSION    AF071538
VERSION      AF071538.1    GI:4007417
KEYWORDS     .
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1894)
AUTHORS      Oettgen,P., Finger,E., Sun,Z., Akbarali,Y., Thamrongsak,U.,
             Boltax,J., Grall,F., Dube,A., Weiss,A., Brown,L., Quinn,G., Kas,K.,
             Endress,G., Kunsch,C. and Libermann,T.A.
TITLE        PDEF, a novel prostate epithelium-specific ets transcription
```

```
Sequencing vector: plasmid; L08752; 98% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-primer Big Dye; 99% of reads
Consensus quality: 276527 bases at least Q40
Consensus quality: 278502 bases at least Q30
Consensus quality: 279605 bases at least Q20
Insert size: 280623; sum-of-contigs
Insert size: 136001; 5.0% error; agarose-fp
Quality coverage: 12.84x in Q20 bases; sum-of-contigs Quality
coverage: 27.11x in Q20 bases; agarose-fp
----------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1   48019: contig of 48019 bp in length
*     48020   48119: gap of     100 bp in length
*     48120   50615: contig of 2496 bp in length
*     50616   50715: gap of     100 bp in length
*     50716  138567: contig of 87852 bp in length
*    138568  138667: gap of     100 bp
*    138668  249115: contig of 110448 bp in length
*    249116  249215: gap of     100 bp
*    249216  252961: contig of 3746 bp in length
*    252962  253061: gap of     100 bp
*    253062  257193: contig of 4132 bp in length
*    257194  257293: gap of     100 bp
*    257294  261324: contig of 4031 bp in length
*    261325  261424: gap of     100 bp
*    261425  267627: contig of 6203 bp in length
*    267628  267727: gap of     100 bp
*    267728  281423: contig of 13696 bp in length.
FEATURES     Location/Qualifiers
source       1..281423
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="6"
             /clone="RP3-377I1"
             /clone_lib="RPCI-3"
misc_feature 1..48019
             /note="assembly_fragment:01145"
misc_feature 48120..50615
             /note="assembly_fragment:03168"
misc_feature 50716..138567
             /note="assembly_fragment:03389"
misc_feature 138668..249115
             /note="assembly_fragment:04917"
misc_feature 249216..252961
             /note="assembly_fragment:06333"
misc_feature 253062..257193
             /note="assembly_fragment:06810"
misc_feature 257294..261324
             /note="assembly_fragment:06900"
misc_feature 261425..267627
             /note="assembly_fragment:07967"
misc_feature 267728..281423
             /note="assembly_fragment:05569"
             clone_end:T7
             vector_side:right"
BASE COUNT   88578 a   58242 c   55761 g   78042 t   800 others
ORIGIN

Query Match      7.6%;      Score 33.6;    DB 78;    Length 281423;
Best Local Similarity    59.4%;    Pred. No. 20;
Matches   57;    Conservative    0;    Mismatches   39;    Indels    0;    Gaps    0;

Qy   135 tcgtccaacatgccctcgagacgtcggtgaagccgtagtttgtacaagactaacaac 194
             || | ||| ||||||   ||||| | |||||||  ||| | |||| || ||||||
```

JOURNAL   factor, interacts with the androgen receptor and activates
           prostate-specific antigen gene expression
           J. Biol. Chem. 275 (2), 1216-1225 (2000)
MEDLINE   20092894
REFERENCE 2 (bases 1 to 1894)
AUTHORS   Libermann,T.A., Oettgen,P., Kunsch,C. and Finger,E.
TITLE   Direct Submission
JOURNAL   Submitted (10-JUN-1998) Medicine, Beth Israel Deaconess Medical
           Center, 330 Brookline Ave., Boston, MA 02215, USA
FEATURES   Location/Qualifiers
     source   1..1894
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6p21.3"
                     /tissue_type="prostate"
     gene   1..1894
                     /gene="PDEF"
     CDS   416..1423
                     /gene="PDEF"
                     /note="prostate epithelium-specific"
                     /codon_start=1
                     /product="Ets transcription factor PDEF"
                     /protein_id="AAC95296.1"
                     /db_xref="GI:4007418"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT   368 a   653 c   571 g   302 t
ORIGIN

Query Match   7.5%; Score 33.4; DB 88; Length 1894;
Best Local Similarity   58.6%; Pred. No. 21;
Matches   58; Conservative   0; Mismatches   41; Indels   0; Gaps   0;

Qy   290 tgctgttaccgtgcagcactcgaagacttcctgctgtactcgctggcagcatagtctg 349
           |||| ||||| | ||||||| | | | ||||| || ||||| | ||||||| ||
Db   443 TGCTCAGACCCGGCTGGGGCTGCGCCCATGCCGCTGCTGTTTGGGCTGGCGGCTGTGTCTG 384

Qy   350 tttgctgctggtctgggacacctgtactacagtac 388
           || |||||||||| | |||| || ||| | ||||
Db   383 TTAGCTGCCTGGTGCCCAGGGAGCTGTCTGCTGCAGTGC 345

RESULT 31
AB031549/c
LOCUS   AB031549   1898 bp   mRNA   PRI   20-JAN-2000
DEFINITION Homo sapiens PSE mRNA for prostate ets, complete cds.
ACCESSION   AB031549
VERSION   AB031549.1   GI:6721497
KEYWORDS   prostate ets.
SOURCE   Homo sapiens adult male prostate epithelium prostate cancer
            cell_line:PC-3 cDNA to mRNA.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (sites)
AUTHORS   Yamada,N., Tamai,Y., Miyamoto,H. and Nozaki,M.
TITLE   Cloning and expression of the mouse Pse gene encoding a novel Ets
            family member
JOURNAL   Gene 241 (2), 267-274 (2000)
MEDLINE   20137504
REFERENCE   2 (bases 1 to 1898)
AUTHORS   Nozaki,M.
TITLE   Direct Submission
JOURNAL   Submitted (25-AUG-1999) to the DDBJ/EMBL/GenBank databases. Masami
            Nozaki, Osaka University, Institute for Microbial Diseases; 3-1,
            Yamadaoka, Suita, Osaka 565-0871, Japan

                     (E-mail:mnozaki@biken.osaka-u.ac.jp, Tel:81-6-879-8338,
                     Fax:81-6-879-8339)
FEATURES   Location/Qualifiers
     source   1..1898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_line="PC-3"
                     /cell_type="prostate cancer"
                     /chromosome="6"
                     /dev_stage="adult"
                     /map="6p21.2-21.3"
                     /sex="male"
                     /tissue_type="prostate epithelium"
                     /note="vector:pUC18"
     gene   406..1413
                     /gene="PSE"
     CDS   406..1413
                     /gene="PSE"
                     /codon_start=1
                     /product="prostate ets"
                     /protein_id="BAA89543.1"
                     /db_xref="GI:6721498"
                     /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                     PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEPEQCPVIDSQAPA
                     GSLDLVPGGLTLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                     WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                     TSPGAIHYCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                     GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                     I"
BASE COUNT   380 a   651 c   567 g   300 t
ORIGIN

Query Match   7.5%; Score 33.4; DB 85; Length 1898;
Best Local Similarity   58.6%; Pred. No. 21;
Matches   58; Conservative   0; Mismatches   41; Indels   0; Gaps   0;

Qy   290 tgctgttaccgtgcagcactcgaagacttcctgctgtactcgctggcagcatagtctg 349
           |||| ||||| | ||||||| | | | ||||| || ||||| || ||||||| ||
Db   433 TGCTCAGACCCGGCTGGGGCTGCGCCCATGCCGCTGCTGTTTGGGCTGGCGGCTGTGTCTG 374

Qy   350 tttgctgctggtctgggacacctgtactacagtac 388
           || |||||||||| | |||| || ||| | ||||
Db   373 TTAGCTGCCTGGTGCCCAGGGAGCTGTCTGCTGCAGTGC 335

RESULT 32
AX048159/c
LOCUS   AX048159   3317 bp   DNA   PAT   15-DEC-2000
DEFINITION Sequence 12 from Patent WO0070092.
ACCESSION   AX048159
VERSION   AX048159.1   GI:11876975
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3317)
AUTHORS   Kaufmann,J. and Xin,H.
TITLE   Expression of ets-domain proteins in cancer
JOURNAL   Patent: WO 0070092-A 12 23-NOV-2000;
            Chiron Corporation (US)
FEATURES   Location/Qualifiers
     source   1..3317
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="GSEF-encoding sequence with promoter (Figs. 2A-2B)"
     promoter   1..1380
                     /note="Promoter"
     CDS   1815..2822
                     /note="GSEF coding sequence"
                     /codon_start=1
                     /protein_id="CAC19004.1"

```
                 /db_xref="GI:11876976"
                 /translation="MGSASPGLSSVSPSHLLLPPDTVSRTGLEKAAAGAVGLERRDWS
                 PSPPATPEQGLSAFYLSYFDMLYPEDSSWAAKAPGASSREEPPEEPEQCPVIDSQAPA
                 GSLDLVPGGIFLEEHSLEQVQSMVVGEVLKDIETACKLLNITADPMDWSPSNVQKWLL
                 WTEHQYRLPPMGKAFQELAGKELCAMSEEQFRQRSPLGGDVLHAHLDIWKSAAWMKER
                 TSPGAIHVCASTSEESWTDSEVDSSCSGQPIHLWQFLKELLLKPHSYGRFIRWLNKEK
                 GIFKIEDSAQVARLWGIRKNRPAMNYDKLSRSIRQYYKKGIIRKPDISQRLVYQFVHP
                 I"
BASE COUNT     710 a   1026 c    970 g    611 t
ORIGIN

Query Match
Best Local Similarity   7.5%;     Score 33.4;   DB 9;   Length 3317;
Matches   58;   Conservative   0;   Mismatches   41;   Indels   0;   Gaps   0;

Qy   290   tgcgtgttacgtgcagcactcgaagacttccttgctgactacgctggcagcatagtctg 349
           ||||   |||  | |    |  |||||||  ||||   |||||  ||||||| || |||| |
Db  1842   TGCTCAGACCCGGGCTGGCCGCCTGGGCGCCATGCCGCTCCTCGTTTGGGCTGGCCGCTGTGTCTG 1783

Qy   350   tttgctgcctggtctggggacacctgtgtactacagtac 388
           || |||||||||| |   |||||||  ||| |||||||  |
Db  1782   TTAGCTGCCTGGTGCCCAGGGAGCTGTGTCTCTGCAGTGC 1744

RESULT   33
AC068504
LOCUS          AC068504     66534 bp     DNA                     HTG        03-MAY-2000
DEFINITION     Homo sapiens chromosome 4 clone RP11-207J21 map 4, LOW-PASS
               SEQUENCE SAMPLING.
ACCESSION      AC068504
VERSION        AC068504.1   GI:7684418
KEYWORDS       HTG; HTGS_PHASE0.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 66534)
AUTHORS        Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE          Homo sapiens chromosome 4, clone RP11-207J21
JOURNAL        Unpublished
REFERENCE      2  (bases 1 to 66534)
AUTHORS        Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
               Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
               Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
               Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
               Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
               Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
               Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
               Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
               Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
               Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
               Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
               McCarthy,M., McEwan,P., McGurk,A., McKernan,A., McPheeters,R.,
               Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
               Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
               O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
               Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
               Roy,A., Santos,R., Stojanovic,N., Subramanian,A., Talamas,J.,
               Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
               Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
               Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
               Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE          Direct Submission
JOURNAL        Submitted (03-MAY-2000) Whitehead Institute/MIT Center for Genome
               Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT        All repeats were identified using RepeatMasker:
               http://ftp.genome.washington.edu/RM/RepeatMasker.html
               ---------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
```

```
               Contact: sequence_submissions@genome.wi.mit.edu
               ----------- Project Information
               Center project name: L10123
               Center clone name: 207_J_21

*  NOTE: This record contains 75 individual
   sequencing reads that have not been assembled into
   contigs. Runs of N are used to separate the reads
   and the order in which they appear is completely
   arbitrary. Low-pass sequence sampling is useful for
   identifying clones that may be gene-rich and allows
   overlap relationships among clones to be deduced.
   However, it should not be assumed that this clone
   will be sequenced to completion. In the event that
   the record is updated, the accession number will
   be preserved.
*         1     778: contig of 778 bp in length
*       779   878: gap of    100 bp
*       879  1660: contig of 782 bp in length
*      1661  1760: gap of    100 bp
*      1761  2542: contig of 782 bp in length
*      2543  2642: gap of    100 bp
*      2643  3416: contig of 774 bp in length
*      3417  3516: gap of    100 bp
*      3517  4294: contig of 778 bp in length
*      4295  4394: gap of    100 bp
*      4395  5194: contig of 800 bp in length
*      5195  5294: gap of    100 bp
*      5295  6093: contig of 799 bp in length
*      6094  6193: gap of    100 bp
*      6194  6985: contig of 792 bp in length
*      6986  7085: gap of    100 bp
*      7086  7897: contig of 812 bp in length
*      7898  7997: gap of    100 bp
*      7998  8767: contig of 770 bp in length
*      8768  8867: gap of    100 bp
*      8868  9648: contig of 781 bp in length
*      9649  9748: gap of    100 bp
*      9749 10533: contig of 785 bp in length
*     10534 10633: gap of    100 bp
*     10634 11395: contig of 762 bp in length
*     11396 11495: gap of    100 bp
*     11496 12270: contig of 775 bp in length
*     12271 12370: gap of    100 bp
*     12371 13151: contig of 781 bp in length
*     13152 13251: gap of    100 bp
*     13252 14045: contig of 794 bp in length
*     14046 14145: gap of    100 bp
*     14146 14930: contig of 785 bp in length
*     14931 15030: gap of    100 bp
*     15031 15825: contig of 795 bp in length
*     15826 15925: gap of    100 bp
*     15926 16706: contig of 781 bp in length
*     16707 16806: gap of    100 bp
*     16807 17595: contig of 789 bp in length
*     17596 17695: gap of    100 bp
*     17696 18486: contig of 791 bp in length
*     18487 18586: gap of    100 bp
*     18587 19361: contig of 775 bp in length
*     19362 19461: gap of    100 bp
*     19462 20243: contig of 782 bp in length
*     20244 20343: gap of    100 bp
*     20344 21126: contig of 783 bp in length
*     21127 21226: gap of    100 bp
*     21227 22013: contig of 787 bp in length
*     22014 22113: gap of    100 bp
*     22114 22888: contig of 775 bp in length
*     22889 22988: gap of    100 bp
*     22989 23775: contig of 787 bp in length
*     23776 23875: gap of    100 bp
*     23876 24674: contig of 799 bp in length
*     24675 24774: gap of    100 bp
*     24775 25565: contig of 791 bp in length
```

```
*   57741    58524: contig of 784 bp in length
*   58525    58624: gap of    100 bp
*   58625    59404: contig of 780 bp in length
*   59405    59504: gap of    100 bp
*   59505    60291: contig of 787 bp in length
*   60292    60391: gap of    100 bp
*   60392    61181: contig of 790 bp in length
*   61182    61281: gap of    100 bp
```

```
Query Match             7.5%;  Score 33.4;  DB 73;  Length 66534;
Best Local Similarity  53.6%;  Pred. No. 23;
Matches   67;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    317 cttccttgctgctactcgctggcagcatagtctgtttgctgcctggtctgggacacctgt  376
          ||| ||  ||| || ||  ||||| ||||| |   | | |||||||| |||| || |||
Db  23886 CTGCCTGCTTGTCGACTCTAAAAGCATCCCCNCTTGCCAACCTGTCTTTGGNAGAAATGT  23945

Qy    377 gtactacagtacttcgttccctttgtagtggtactacttcttagccatgtactactgtc  436
          ||   || || |||  ||| || |  ||||  ||  | |||| |||| |||
Db  23946 CTATTCAGGTTCATTGCTCATTTTAAATTGGGTTGTTTTGTTTGCCATTGAGTTGTGTG  24005

Qy    437 tgttc  441
          |||||
Db  24006 AGTTC  24010
```

```
RESULT 34
AC022781/c
LOCUS       AC022781     200729 bp    DNA            HTG        26-SEP-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-199H17 map 11, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC022781
VERSION     AC022781.5  GI:10305259
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200729)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-199H17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200729)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkly,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 26, 2000 this sequence version replaced gi:9154653.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
```

```
*   25566    25665: gap of    100 bp
*   25666    26451: contig of 786 bp in length
*   26452    26551: gap of    100 bp
*   26552    27347: contig of 796 bp in length
*   27348    27447: gap of    100 bp
*   27448    28248: contig of 801 bp in length
*   28249    28348: gap of    100 bp
*   28349    29135: contig of 787 bp in length
*   29136    29235: gap of    100 bp
*   29236    30021: contig of 786 bp in length
*   30022    30121: gap of    100 bp
*   30122    30921: contig of 800 bp in length
*   30922    31021: gap of    100 bp
*   31022    31829: contig of 808 bp in length
*   31830    31929: gap of    100 bp
*   31930    32720: contig of 791 bp in length
*   32721    32820: gap of    100 bp
*   32821    33611: contig of 791 bp in length
*   33612    33711: gap of    100 bp
*   33712    34502: contig of 791 bp in length
*   34503    34602: gap of    100 bp
*   34603    35374: contig of 772 bp in length
*   35375    35474: gap of    100 bp
*   35475    36258: contig of 784 bp in length
*   36259    36358: gap of    100 bp
*   36359    37175: contig of 817 bp in length
*   37176    37275: gap of    100 bp
*   37276    38069: contig of 794 bp in length
*   38070    38169: gap of    100 bp
*   38170    38945: contig of 776 bp in length
*   38946    39045: gap of    100 bp
*   39046    39836: contig of 791 bp in length
*   39837    39936: gap of    100 bp
*   39937    40746: contig of 810 bp in length
*   40747    40846: gap of    100 bp
*   40847    41641: contig of 795 bp in length
*   41642    41741: gap of    100 bp
*   41742    42535: contig of 794 bp in length
*   42536    42635: gap of    100 bp
*   42636    43402: contig of 767 bp in length
*   43403    43502: gap of    100 bp
*   43503    44281: contig of 779 bp in length
*   44282    44381: gap of    100 bp
*   44382    45156: contig of 775 bp in length
*   45157    45256: gap of    100 bp
*   45257    46041: contig of 785 bp in length
*   46042    46141: gap of    100 bp
*   46142    46940: contig of 799 bp in length
*   46941    47040: gap of    100 bp
*   47041    47834: contig of 794 bp in length
*   47835    47934: gap of    100 bp
*   47935    48721: contig of 787 bp in length
*   48722    48821: gap of    100 bp
*   48822    49609: contig of 788 bp in length
*   49610    49709: gap of    100 bp
*   49710    50509: contig of 800 bp in length
*   50510    50609: gap of    100 bp
*   50610    51392: contig of 783 bp in length
*   51393    51492: gap of    100 bp
*   51493    52284: contig of 792 bp in length
*   52285    52384: gap of    100 bp
*   52385    53173: contig of 789 bp in length
*   53174    53273: gap of    100 bp
*   53274    54038: contig of 765 bp in length
*   54039    54138: gap of    100 bp
*   54139    54931: contig of 793 bp in length
*   54932    55031: gap of    100 bp
*   55032    55826: contig of 795 bp in length
*   55827    55926: gap of    100 bp
*   55927    56742: contig of 816 bp in length
*   56743    56842: gap of    100 bp
*   56843    57640: contig of 798 bp in length
*   57641    57740: gap of    100 bp
```

```
Center project name: L5913
Center clone name: 199_H_17
-------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180464 bases at least Q40
Consensus quality: 190837 bases at least Q30
Consensus quality: 194945 bases at least Q20
Insert size: 206000; agarose-fp
Insert size: 197129; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs
-------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    7021: contig of  7021 bp in length
*     7022    7121: gap of      100 bp
*     7122    8310: contig of  1189 bp in length
*     8311    8410: gap of      100 bp
*     8411    9411: contig of  1001 bp in length
*     9412    9511: gap of      100 bp
*     9512   11581: contig of  2070 bp in length
*    11582   11681: gap of      100 bp
*    11682   13342: contig of  1661 bp in length
*    13343   13442: gap of      100 bp
*    13443   16024: contig of  2582 bp in length
*    16025   16124: gap of      100 bp
*    16125   17440: contig of  1316 bp in length
*    17441   17540: gap of      100 bp
*    17541   18922: contig of  1382 bp in length
*    18923   19022: gap of      100 bp
*    19023   21353: contig of  2331 bp in length
*    21354   21453: gap of      100 bp
*    21454   23674: contig of  2221 bp in length
*    23675   23774: gap of      100 bp
*    23775   25868: contig of  2094 bp in length
*    25869   25968: gap of      100 bp
*    25969   27966: contig of  1998 bp in length
*    27967   28066: gap of      100 bp
*    28067   31051: contig of  2985 bp in length
*    31052   31151: gap of      100 bp
*    31152   33236: contig of  2085 bp in length
*    33237   33336: gap of      100 bp
*    33337   36285: contig of  2949 bp in length
*    36286   36385: gap of      100 bp
*    36386   39476: contig of  3091 bp in length
*    39477   39576: gap of      100 bp
*    39577   42584: contig of  3008 bp in length
*    42585   42684: gap of      100 bp
*    42685   46164: contig of  3480 bp in length
*    46165   46264: gap of      100 bp
*    46265   49973: contig of  3709 bp in length
*    49974   50073: gap of      100 bp
*    50074   55389: contig of  5316 bp in length
*    55390   55489: gap of      100 bp
*    55490   59283: contig of  3794 bp in length
*    59284   59383: gap of      100 bp
*    59384   62877: contig of  3494 bp in length
*    62878   62977: gap of      100 bp
*    62978   68182: contig of  5205 bp in length
*    68183   68282: gap of      100 bp
*    68283   73862: contig of  5580 bp in length
*    73863   73962: gap of      100 bp
*    73963   83038: contig of  9076 bp in length
*    83039   83138: gap of      100 bp
*    83139   87822: contig of  4684 bp in length
*    87823   87922: gap of      100 bp
*    87923   92086: contig of  4164 bp in length
*    92087   92186: gap of      100 bp
*    92187   97007: contig of  4821 bp in length
*    97008   97107: gap of      100 bp
*    97108  123018: contig of 25911 bp in length
*   123019  123118: gap of      100 bp
*   123119  129307: contig of  6189 bp in length
*   129308  129407: gap of      100 bp
*   129408  137163: contig of  7756 bp in length
*   137164  137263: gap of      100 bp
*   137264  147510: contig of 10247 bp in length
*   147511  147610: gap of      100 bp
*   147611  155502: contig of  7892 bp in length
*   155503  155602: gap of      100 bp
*   155603  168939: contig of 13337 bp in length
*   168940  169039: gap of      100 bp
*   169040  181336: contig of 12297 bp in length
*   181337  181436: gap of      100 bp
*   181437  195395: contig of 13359 bp in length
*   195396  195495: gap of      100 bp
*   195496  200729: contig of  5234 bp in length.

FEATURES             Location/Qualifiers
     source          1..200729
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /map="11"
                     /clone="RP23-199H17"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1..7021
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    7122..8310
                     /note="assembly_fragment"
     misc_feature    8411..9411
                     /note="assembly_fragment"
     misc_feature    9512..11581
                     /note="assembly_fragment"
     misc_feature    11682..13342
                     /note="assembly_fragment"
     misc_feature    13443..16024
                     /note="assembly_fragment"
     misc_feature    16125..17440
                     /note="assembly_fragment"
     misc_feature    17541..18922
                     /note="assembly_fragment"
     misc_feature    19023..21353
                     /note="assembly_fragment"
     misc_feature    21454..23674
                     /note="assembly_fragment"
     misc_feature    23775..25868
                     /note="assembly_fragment"
     misc_feature    25969..27966
                     /note="assembly_fragment"
     misc_feature    28067..31051
                     /note="assembly_fragment"
     misc_feature    31152..33236
                     /note="assembly_fragment"
     misc_feature    33337..36285
                     /note="assembly_fragment"
     misc_feature    36386..39476
                     /note="assembly_fragment"
     misc_feature    39577..42584
                     /note="assembly_fragment"
     misc_feature    42685..46164
                     /note="assembly_fragment"
     misc_feature    46265..49973
                     /note="assembly_fragment"
     misc_feature    50074..55389
                     /note="assembly_fragment"
     misc_feature    55490..59283
                     /note="assembly_fragment"
```

Mon Sep 10 07:42:20 2001

```
misc_feature          /note="assembly_fragment"
                  59384..62877
                      /note="assembly_fragment"
misc_feature          62978..68182
                      /note="assembly_fragment"
misc_feature          68283..73862
                      /note="assembly_fragment"
misc_feature          73963..83038
                      /note="assembly_fragment"

Query Match       7.5%;   Score 33.4;  DB 67;  Length 200729;
Best Local Similarity  47.4%;  Pred. No. 23;
Matches 100;  Conservative   0;  Mismatches  111;  Indels   0;  Gaps     0;

Qy     231 ctgtttactgtgtagccgcttgatgacttccttgctgtactccgcttgagcgccatttgagt 290
               ||| ||||||||||||| ||||| ||| || || |||  ||| ||||| |||||
Db  192888 CTCTGAACAGGGGCACTTCCTTCTTCTTACTTACTTCTTTTTTAATCTTTCTCTTGAAA 192829

Qy     291 gctgtttaccgtgcagcactcgaagacttccttgctgtactcgctggcagcatagtctgt 350
               | ||||| ||||| || ||| ||  |||| | ||||||  | ||| ||
Db  192828 AAGGCTCTCTATGTAGCCCTGGATATCCTACTCATGTAGATCAGGTTTACCCAAAACCCT 192769

Qy     351 ttgctgccctggtctggggacacctgtgtactacagtacttcccttgtagtggta 410
               ||||||| ||||| || ||| || ||| |||||| | |||||| ||||| |||
Db  192768 TTGCTTCCCCACGCTGGATCCAAAGGTGCAACTGTACCTGGCTGCTTCTGCTTCTT 192709

Qy     411 ctacttcttagcatgtcactcgtgtctgttc 441
               || ||||||| || |||  || ||| |||||
Db  192708 CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC 192678

RESULT  35
AL356364/c
LOCUS       AL356364    128819 bp    DNA               HTG      15-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-328I22, *** SEQUENCING IN
            PROGRESS ***, 27 unordered pieces.
ACCESSION   AL356364
VERSION     AL356364.5    GI:12539661
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 128219)
AUTHORS     Plumb,B.
TITLE       Direct Submission
JOURNAL     Submitted (14-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Jan 26, 2001 this sequence version replaced gi:9797506.
            ---------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ---------- Project Information
            Center project name: bA328I22
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 113756 bases at least Q40
            Consensus quality: 118150 bases at least Q30
            Consensus quality: 120849 bases at least Q20
            Insert size: 125619; sum-of-contigs
            Insert size: 178864; 11.7% error; agarose-fp
            Quality coverage: 2.55x in Q20 bases; sum-of-contigs Quality
            coverage: 2.11x in Q20 bases; agarose-fp
            -------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*            1      4111: contig of 4111 bp in length
*         4112      4211: gap of          100 bp
*         4212     11454: contig of 7243 bp in length
*        11455     11554: gap of          100 bp
*        11555     16224: contig of 4670 bp in length
*        16225     16324: gap of          100 bp
*        16325     19960: contig of 3636 bp in length
*        19961     20060: gap of          100 bp
*        20061     24131: contig of 4071 bp in length
*        24132     24231: gap of          100 bp
*        24232     28443: contig of 4212 bp in length
*        28444     28543: gap of          100 bp
*        28544     31045: contig of 2502 bp in length
*        31046     31145: gap of          100 bp
*        31146     36931: contig of 5786 bp in length
*        36932     37031: gap of          100 bp
*        37032     42061: contig of 5030 bp in length
*        42062     42161: gap of          100 bp
*        42162     44625: contig of 2464 bp in length
*        44626     44725: gap of          100 bp
*        44726     47635: contig of 2910 bp in length
*        47636     47735: gap of          100 bp
*        47736     51054: contig of 3319 bp in length
*        51055     51154: gap of          100 bp
*        51155     57015: contig of 5861 bp in length
*        57016     57115: gap of          100 bp
*        57116     64974: contig of 7859 bp in length
*        64975     65074: gap of          100 bp
*        65075     71197: contig of 6123 bp in length
*        71198     71297: gap of          100 bp
*        71298     74488: contig of 3191 bp in length
*        74489     74588: gap of          100 bp
*        74589     79354: contig of 4766 bp in length
*        79355     79454: gap of          100 bp
*        79455     85107: contig of 5653 bp in length
*        85108     85207: gap of          100 bp
*        85208     89243: contig of 4036 bp in length
*        89244     89343: gap of          100 bp
*        89344     92074: contig of 2731 bp in length
*        92075     92174: gap of          100 bp
*        92175     99269: contig of 7095 bp in length
*        99270     99369: gap of          100 bp
*        99370    103711: contig of 4342 bp in length
*       103712    103811: gap of          100 bp
*       103812    106764: contig of 2953 bp in length
*       106765    106864: gap of          100 bp
*       106865    113063: contig of 6199 bp in length
*       113064    113163: gap of          100 bp
*       113164    118481: contig of 5318 bp in length
*       118482    118581: gap of          100 bp
*       118582    122075: contig of 3494 bp in length
*       122076    122175: gap of          100 bp
*       122176    128219: contig of 6044 bp in length

FEATURES             Location/Qualifiers
     source          1..128219
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-328I22"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..4111
                     /note="assembly_fragment:00284
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    4212..11454
                     /note="assembly_fragment:01334
                     fragment_chain:1"
     misc_feature    11555..16224
```

```
                                      Db  91890  TGTCCTTCTTCGCTAGAGGAGTCCCTTGCTGTGTGATTCTTCTTGGATACATTTAGGACT 91831
                                                 |  | || |   ||| |  ||| || | | | | | |||| | | | |||||  |||||| || |  ||
                                      Qy    294  gttaccgtg 303
                                                 | | |||||
                                      Db  91830  TTCTTCCGTG 91821

 misc_feature    /note="assembly_fragment:00579
                 fragment_chain:1"
                 16325..19960
 misc_feature    /note="assembly_fragment:00196    RESULT 36
                 fragment_chain:2"                 AL513283
                 2006..24131                       LOCUS       AL513283    180381 bp    DNA              HTG       07-FEB-2001
 misc_feature    /note="assembly_fragment:00024    DEFINITION  Homo sapiens chromosome 1 clone RP11-95P13, *** SEQUENCING IN
                 fragment_chain:2"                             PROGRESS ***, 17 unordered pieces.
                 24232..28443                      ACCESSION   AL513283
 misc_feature    /note="assembly_fragment:01340    VERSION     AL513283.3  GI:12718173
                 fragment_chain:2"                 KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
                 28544..31045                      SOURCE      human.
 misc_feature    /note="assembly_fragment:00423      ORGANISM  Homo sapiens
                 fragment_chain:3"                             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                 31146..36931                                  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 misc_feature    /note="assembly_fragment:01164    REFERENCE   1  (bases 1 to 180381)
                 fragment_chain:3"                 AUTHORS     Pavitt,R.
                 37032..42061                      TITLE       Direct Submission
 misc_feature    /note="assembly_fragment:00612    JOURNAL     Submitted (06-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
                 fragment_chain:3"                             CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
                 42162..44625                                  requests: clonerequest@sanger.ac.uk
 misc_feature    /note="assembly_fragment:00259    COMMENT     On Feb 8, 2001 this sequence version replaced gi:12619151.
                 fragment_chain:4"                             -------------- Genome Center
                 44726..47635                                  Center: Sanger Centre
 misc_feature    /note="assembly_fragment:01001                Center code: SC
                 fragment_chain:4"                             Web site: http://www.sanger.ac.uk
                 47736..51054                                  Contact: humquery@sanger.ac.uk
 misc_feature    /note="assembly_fragment:00296                -------------- Project Information
                 fragment_chain:5"                             Center project name: bA95P13
                 51155..57015                                  -------------- Summary Statistics
 misc_feature    /note="assembly_fragment:00190                Assembly program: XGAP4; version 4.5
                 fragmen_chain:5"                              Sequencing vector: plasmid; L08752; 100% of reads
                 57116..64974                                  Chemistry: Dye-terminator Big Dye; 100% of reads
 misc_feature    /note="assembly_fragment:00875                Consensus quality: 172799 bases at least Q40
                 fragmn_chain:6"                               Consensus quality: 175859 bases at least Q30
                 65075..71197                                  Consensus quality: 177452 bases at least Q20
 misc_feature    /note="assembly_fragment:00537                Insert size: 178781; sum-of-contigs
                 fragment_chain:6"                             Insert size: 175883; 17.3% error; agarose-fp
                 71299..74488                                  Quality coverage: 4.04x in Q20 bases; sum-of-contigs Quality
 misc_feature    /note="assembly_fragment:00887                coverage: 4.32x in Q20 bases; agarose-fp
                 fragment_chain:7"                             --------------
                 74589..79354                      *     NOTE: This is a 'working draft' sequence. It currently
 misc_feature    /note="assembly_fragment:00909    *     consists of 17 contigs. The true order of the pieces
                 fragment_chain:7"                 *     is not known and their order in this sequence record is
                 79455..85107                      *     arbitrary. Gaps between the contigs are represented as
 misc_feature    /note="assembly_fragment:00022"   *     runs of N, but the exact sizes of the gaps are unknown.
                 85208..89243                      *     This record will be updated with the finished sequence
 misc_feature    /note="assembly_fragment:00118    *     as soon as it is available and the accession number will
                 89344..92074                      *     be preserved.
 misc_feature    /note="assembly_fragment:00276"   *          1   11197: contig of 11197 bp in length
                 92175..99269                      *      11198   11297: gap of    100 bp
 misc_feature    /note="assembly_fragment:00380    *      11298   54629: contig of 43332 bp in length
                 99370..103711                     *      54630   54729: gap of    100 bp
 misc_feature    /note="assembly_fragment:00586"   *      54730   65732: contig of 11003 bp in length
                 103812..106764                    *      65733   65832: gap of    100 bp
 misc_feature    /note="assembly_fragment:00685"   *      65833   69227: contig of 3395 bp in length
                 106865..113063                    *      69228   69327: gap of    100 bp
 misc_feature    /note="assembly_fragment:00693"   *      69328   72585: contig of 3258 bp in length
                 113164..118481                    *      72586   72685: gap of    100 bp
 misc_feature    /note="assembly_fragment:00739"   *      72686   77520: contig of 4835 bp in length
                 118582..122075                    *      77521   77620: gap of    100 bp
 misc_feature    /note="assembly_fragment:00905"   *      77621   87310: contig of 9690 bp in length
                 122176..128219                    *      87311   87410: gap of    100 bp
 misc_feature    /note="assembly_fragment:01254"   *      87411   90165: contig of 2755 bp in length
                                                   *      90166   90265: gap of    100 bp
 BASE COUNT    38763 a  24884 c  24158 g  37809 t  2605 others   90266  130408: contig of 40143 bp in length
 ORIGIN                                            *     130409  130508: gap of    100 bp
                                                   *     130509  132537: contig of 2029 bp in length

 Query Match         7.5%;   Score 33.2;  DB 80;  Length 128219;
 Best Local Similarity  67.1%;    Pred. No. 26;
 Matches   47;  Conservative   0;  Mismatches   23;  Indels   0;  Gaps   0;

 Qy   234  tttactgtgtgagcgcttgatgactccttgctgtactccgcttggagccattgagtgct 293
```

```
ORIGIN

    Query Match                7.5%;    Score 33.2;    DB 81;    Length 180381;
    Best Local Similarity     67.1%;    Pred. No. 27;
    Matches    47;    Conservative    0;    Mismatches    23;    Indels    0;    Gaps    0;

    Qy    234 ttactgtgtagcgcttgatgacttcctgctgtactccgcttgagccattgagtgct 293
                ||||||| ||||||| ||||||||||||||||||||||||| || ||| |||| ||
    Db 142733 TGTCCTCTTCTTCGCTAGAGAGTCCCTTGCTGTGATTCTCTTGGATACATTTTAGGACT 142792

    Qy    294 gtttaccgtg 303
                || |||||
    Db 142793 TTCTTCCGTG 142802

RESULT 37
AC023785/c
LOCUS         AC023785           209599 bp    DNA        HTG        07-JUL-2000
DEFINITION    Homo sapiens chromosome 6 clone RP11-507C10, WORKING DRAFT
              SEQUENCE, 18 unordered pieces.
ACCESSION     AC023785
VERSION       AC023785.3  GI:8099307
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 209599)
  AUTHORS     Waterston,R.H.
  TITLE       The sequence of Homo sapiens clone
  JOURNAL     Unpublished
REFERENCE     2  (bases 1 to 209599)
  AUTHORS     Waterston,R.H.
  TITLE       Direct Submission
  JOURNAL     Submitted (17-FEB-2000) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT       On May 27, 2000 this sequence version replaced gi:7109627.

              ----------------- Genome Center ----------------
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site:http://genome.wustl.edu/gsc/index.shtml
              ------------- Project Information --------------
              Center project name: H_NH0507C10
              --------------- Summary Statistics -------------
              Sequencing vector: M13; 78%
              Sequencing vector: plasmid; 22%
              Chemistry: Dye-primer ET; 78% of reads
              Chemistry: Dye-terminator Big Dye; 22% of reads
              Assembly program: Phrap; version 0.990319
              Consensus quality: 199546 bases at least Q40
              Consensus quality: 202608 bases at least Q30
              Consensus quality: 204160 bases at least Q20
              Insert size: 210000; agarose-fp
              Insert size: 207899; sum-of-contigs
              Quality coverage: 6.01 in Q20 bases; agarose-fp
              Quality coverage: 6.11 in Q20 bases; sum-of-contigs

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1      1478: contig of 1478 bp in length
            *      1479      1578: gap of unknown length
            *      1579      3500: contig of 1922 bp in length
            *      3501      3600: gap of unknown length
            *      3601      6635: contig of 3035 bp in length
```

```
            * 132538  132637: gap of     100 bp
            * 132638  138569: contig of 5932 bp in length
            * 138570  138669: gap of     100 bp
            * 138670  141380: contig of 2711 bp in length
            * 141381  141480: gap of     100 bp
            * 141481  151559: contig of 10079 bp in length
            * 151560  151659: gap of     100 bp
            * 151660  154357: contig of 2698 bp in length
            * 154358  154457: gap of     100 bp
            * 154458  162154: contig of 7697 bp in length
            * 162155  162254: gap of     100 bp
            * 162255  170186: contig of 7932 bp in length
            * 170187  170286: gap of     100 bp
            * 170287  180381: contig of 10095 bp in length.

FEATURES             Location/Qualifiers
     source          1..180381
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-95P13"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..1197
                     /note="assembly_fragment:00204
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    11298..54629
                     /note="assembly_fragment:00098
                     fragment_chain:1"
     misc_feature    54730..65732
                     /note="assembly_fragment:00839
                     fragment_chain:2"
     misc_feature    65833..69227
                     /note="assembly_fragment:01812
                     fragment_chain:2"
     misc_feature    69328..72585
                     /note="assembly_fragment:00711
                     fragment_chain:2"
     misc_feature    72686..77520
                     /note="assembly_fragment:01336
                     fragment_chain:2"
     misc_feature    77621..87310
                     /note="assembly_fragment:01246
                     fragment_chain:2"
     misc_feature    87411..90165
                     /note="assembly_fragment:01290
                     fragment_chain:2"
     misc_feature    90266..130408
                     /note="assembly_fragment:00070
                     fragment_chain:2"
     misc_feature    130509..132537
                     /note="assembly_fragment:01353
                     fragment_chain:2"
     misc_feature    132638..138569
                     /note="assembly_fragment:00925
                     fragment_chain:3"
     misc_feature    138670..141380
                     /note="assembly_fragment:01044
                     fragment_chain:3"
     misc_feature    141481..151559
                     /note="assembly_fragment:01420
                     fragment_chain:3"
     misc_feature    151660..154357
                     /note="assembly_fragment:01190
                     fragment_chain:4"
     misc_feature    154458..162154
                     /note="assembly_fragment:01090
                     fragment_chain:4"
     misc_feature    162255..170186
                     /note="assembly_fragment:00648"
     misc_feature    170287..180381
                     /note="assembly_fragment:01673"
BASE COUNT   53764 a  34451 c  34528 g  56033 t   1605 others
```

```
COMMENT                    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
                           requests: clonerequest@sanger.ac.uk
                           On Apr 17, 2001 this sequence version replaced gi:13620456.
                           -------------------- Genome Center
                           Center: Sanger Centre
                           Center code: SC
                           Web site: http://www.sanger.ac.uk
                           Contact: humquery@sanger.ac.uk
                           -------------------- Project Information
                           Center project name: bA507C10
                           -------------------- Summary Statistics
                           Assembly program: XGAP4; version 4.5
                           Sequencing vector: plasmid; L08752; 100% of reads
                           Chemistry: Dye-terminator Big Dye; 100% of reads
                           Consensus quality: 215043 bases at least Q40
                           Consensus quality: 215569 bases at least Q30
                           Consensus quality: 215877 bases at least Q20
                           Insert size: 216160; sum-of-contigs
                           Insert size: 219388; 1.5% error; agarose-fp
                           Quality coverage: 7.69x in Q20 bases; sum-of-contigs Quality
                           coverage: 7.70x in Q20 bases; agarose-fp
                         * NOTE: This is a 'working draft' sequence. It currently
                         * consists of 10 contigs. The true order of the pieces
                         * is not known and their order in this sequence record is
                         * arbitrary. Gaps between the contigs are represented as
                         * runs of N, but the exact sizes of the gaps are unknown.
                         * This record will be updated with the finished sequence
                         * as soon as it is available and the accession number will
                         * be preserved.
                         *       1   36980: contig of 36980 bp in length
                         *   36981   37080: gap of    100 bp
                         *   37081   48800: contig of 11721 bp in length
                         *   48801   48901: gap of    100 bp
                         *   48902   55638: contig of  6737 bp in length
                         *   55639   55738: gap of    100 bp
                         *   55739   64049: contig of  8311 bp in length
                         *   64050   64149: gap of    100 bp
                         *   64150  143890: contig of 79741 bp in length
                         *  143891  143990: gap of    100 bp
                         *  143991  157506: contig of 13516 bp in length
                         *  157507  157606: gap of    100 bp
                         *  157607  182957: contig of 25351 bp in length
                         *  182958  183057: gap of    100 bp
                         *  183058  185348: contig of  2291 bp in length
                         *  185349  185448: gap of    100 bp
                         *  185449  192902: contig of  7454 bp in length
                         *  192903  193002: gap of    100 bp
                         *  193003  217060: contig of 24058 bp in length.
FEATURES             Location/Qualifiers
     source          1..217060
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q25.2-26"
                     /clone="RP11-507C10"
                     /clone_lib="RPCI-11.2"
     misc_feature    1..36980
                     /note="assembly_fragment:02266
                     fragment_chain:1"
     misc_feature    37081..48801
                     /note="assembly_fragment:01215
                     fragment_chain:1"
     misc_feature    48902..55638
                     /note="assembly_fragment:00484
                     fragment_chain:1"
     misc_feature    55739..64049
                     /note="assembly_fragment:00882
                     fragment_chain:1"
     misc_feature    64150..143890
                     /note="assembly_fragment:00373
                     fragment_chain:1"
     misc_feature    143991..157506
```

```
     *       6636   6735: gap of unknown length
     *       6736   9604: contig of 2869 bp in length
     *       9605   9704: gap of unknown length
     *       9705  12323: contig of 2619 bp in length
     *      12324  12423: gap of unknown length
     *      12424  15033: contig of 2610 bp in length
     *      15034  15133: gap of unknown length
     *      15134  19732: contig of 4599 bp in length
     *      19733  19832: gap of unknown length
     *      19833  26562: contig of 6730 bp in length
     *      26563  26662: gap of unknown length
     *      26663  32701: contig of 6039 bp in length
     *      32702  32801: gap of unknown length
     *      32802  41794: contig of 8993 bp in length
     *      41795  41894: gap of unknown length
     *      41895  52103: contig of 10209 bp in length
     *      52104  52203: gap of unknown length
     *      52204  61862: contig of 9659 bp in length
     *      61863  61962: gap of unknown length
     *      61963  73398: contig of 11436 bp in length
     *      73399  73498: gap of unknown length
     *      73499  85215: contig of 11717 bp in length
     *      85216  85315: gap of unknown length
     *      85316  99877: contig of 14562 bp in length
     *      99878  99977: gap of unknown length
     *      99978 109794: contig of 9817 bp in length
     *     109795 109894: gap of unknown length
     *     109895 158727: contig of 48733 bp in length
     *     158728 158827: gap of unknown length
     *     158828 209599: contig of 50872 bp in length.
FEATURES             Location/Qualifiers
     source          1..209599
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-507C10"
BASE COUNT     55365 a  47428 c  48595 g  56476 t   1735 others
ORIGIN

Query Match         7.4%;  Score 32.8;  DB 68;  Length 209599;
Best Local Similarity  54.0%;  Pred. No. 35;
Matches  67; Conservative   0; Mismatches   57; Indels   0; Gaps   0;

Qy     43 atctgaaggaaggacaaagacacctccaagccggaccggacacggctactgtagtcggcacggc 102
          | |||||| ||| || |||| |||  |||| || ||| ||  |||| |||| |||||||||
Db  64349 AGCAGAAGGCAGGGAGGAACACACATCTGGCCTGCCACACGTCTCATCTGGCTTGCAGAGG 64290

Qy    103 ataatcggaaaacagtccctgccctttgagggtcgtctcaaacatgcccctcgagacgtc 162
          || ||||||| |  |||| |||  |||||| ||||  |||  ||| ||| | |||
Db  64289 ATCTTGGAAAAAACTGTAACTCATTGCCAATATATTTAAGAACTGGGAATTGGGCACGGT 64230

Qy    163 ggtg 166
          | ||
Db  64229 GGCG 64226

RESULT 38
AL589931/c
LOCUS       AL589931   217060 bp   DNA                 HTG       14-APR-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-507C10 map q25.2-26, ***
            SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION   AL589931
VERSION     AL589931.6  GI:13660874
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 217060)
AUTHORS     Sehra,H.
TITLE       Direct Submission
JOURNAL     Submitted (13-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
```

```
     misc_feature         /note="assembly_fragment:00039
                          fragment_chain:1"
                          157607..182957
     misc_feature         /note="assembly_fragment:00361
                          fragment_chain:1"
                          183058..185348
     misc_feature         /note="assembly_fragment:02093
                          fragment_chain:1"
                          185449..192902
     misc_feature         /note="assembly_fragment:00794
                          fragment_chain:1"
                          193003..217060
     misc_feature         /note="assembly_fragment:01647
                          fragment_chain:1"
BASE COUNT   57477 a  50328 c  49161 g  59193 t    901 others
ORIGIN

Query Match           7.4%;   Score 32.8;  DB 82;  Length 217060;
Best Local Similarity 54.0%;  Pred. No. 35;
Matches  67; Conservative  0; Mismatches  57; Indels  0; Gaps  0;

Qy     43 atctgaaggaaggacaaagacacctccaagccgacacggctactgtagctggcacggc 102
           |||||||||||||||| ||||||| | ||||| | ||||||||||||||||||||||
Db 195184 AGCAGAAGGCAGGGAGGAACACCACATCTGGCCTGCACACGTCTCATCTGGCTTGCAGAGG 195125

Qy    103 ataatgaaaacagtccctgcccttttgagggtcgtcaaacatgcccctgagacgtc 162
           || |||||||    |||||||||   |||||||||  ||||||| |   || |||
Db 195124 ATCTTGGAAAAAAACTGTAACTCATTTGCCAATATTTAAGAACTGGGAATTGGGCACCGT 195065

Qy    163 gtg 166
           |||
Db 195064 GGGG 195061

RESULT 39
AC023681
LOCUS       AC023681   157890 bp    DNA        HTG       29-MAY-2000
DEFINITION  Drosophila melanogaster clone RPCI98-39P3, WORKING DRAFT SEQUENCE,
            8 unordered pieces.
ACCESSION   AC023681
VERSION     AC023681.3  GI:8101328
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      fruit fly.
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 157890)
AUTHORS     Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Donah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
            Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
            Holloway,C., Hosak.H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
            Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
            Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozado,R.J., Lu,J.,
            Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
            Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
            Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
            Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
            Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R.,
            Tabor,P., Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wahbah,M.,
            Watlington,S., Weinstock,G., Weinstock,I.R., Williamson,A.,
            Worley,K., Wren,J., Wrensford,G., Yu,W., Zhou,X., Nelson,D. and
            Gibbs,R.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 157890)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 29, 2000 this sequence version replaced gi:6997251.
            --------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            --------------- Project Information
              Center project name: DRHR
              Center clone name: RPCI98-39P3
            --------------- Summary Statistics
              Sequencing vector: M13; L08821
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.980611
              Consensus quality: 148728 bases at least Q40
              Consensus quality: 154568 bases at least Q30
              Consensus quality: 157627 bases at least Q20
              Estimated insert size: 133000; agarose-fp estimation
              Estimated insert size: 157182; sum-of-contigs estimation
              Quality coverage: 5.3x in Q20 bases; agarose-fp estimation
              Quality coverage: 4.5x in Q20 bases; sum-of-contigs estimation
            ---------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * ---------------
            *        1     47574: contig of 47574 bp in length
            *    47575     47674: gap of unknown length
            *    47675     79405: contig of 31731 bp in length
            *    79406     79505: gap of unknown length
            *    79506    107003: contig of 27498 bp in length
            *   107004    107103: gap of unknown length
            *   107104    133489: contig of 26386 bp in length
            *   133490    133589: gap of unknown length
            *   133590    153613: contig of 20024 bp in length
            *   153614    153713: gap of unknown length
            *   153714    155467: contig of 1754 bp in length
            *   155468    155567: gap of unknown length
            *   155568    156667: contig of 1100 bp in length
            *   156668    156767: gap of unknown length
            *   156768    157890: contig of 1123 bp in length.
FEATURES             Location/Qualifiers
     source          1..157890
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="RPCI98-39P3"
     misc_feature    60176..60275
                     /note="assembly_fragment:contig16;
                     vector:Vector_pBACe3.6;
                     vector_side:left"
     misc_feature    complement(87790..87889)
                     /note="assembly_fragment:contig17;
                     vector:Vector_pBACe3.6;
                     vector_side:right"
BASE COUNT   45713 a  32660 c  32796 g  46003 t    718 others
ORIGIN

Query Match           7.3%;   Score 32.6;  DB 68;  Length 157890;
Best Local Similarity 58.9%;  Pred. No. 41;
Matches  56; Conservative  0; Mismatches  39; Indels  0; Gaps  0;

Qy    209 tgttatcggagccattcgtgtcgtgttttactgtgtagtagccgcttccttgctgta 268
           |||| ||  ||  |||||| ||| || |||||||| |||| ||||||||||| || ||
Db  10445 TGTTGTCCTTGTGTTGTGTTTTTACTGTGTTATTGTGTTGTCCTTGTGTTGTTTGTTTGTG 10504

Qy    269 ctccgcttggagccattgagtgctgtttaccgtg 303
```

```
Db  10505  TTGTCCTTGTTGTTTGTTTGTTGTTGTTGTCCTTG 10539
                |  |||| |  |  | |  ||  | ||||  || ||

RESULT 40
AL137002/c
LOCUS       AL137002    177407 bp    DNA              HTG       10-MAR-2001
DEFINITION  Homo sapiens chromosome 13 clone RP11-98F14 map q22.1-31.1, ***
            SEQUENCING IN PROGRESS ***, 7 unordered pieces.
ACCESSION   AL137002
VERSION     AL137002.18  GI:13274222
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177407)
AUTHORS     Bates,K.
TITLE       Direct Submission
JOURNAL     Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Mar 11, 2001 this sequence version replaced gi:13184282.
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: bA30E17
            -------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 3% of reads
            Sequencing vector: plasmid; L08752; 96% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 173225 bases at least Q40
            Consensus quality: 175134 bases at least Q30
            Consensus quality: 176014 bases at least Q20
            Insert size: 176807; sum-of-contigs
            Insert size: 175795; 5.9% error; agarose-fp
            Quality coverage: 9.58x in Q20 bases; sum-of-contigs Quality
            coverage: 11.74x in Q20 bases; agarose-fp
            --------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2152: contig of 2152 bp in length
            *     2153   2252: gap of       100 bp
            *     2253   5329: contig of 3077 bp in length
            *     5330   5429: gap of       100 bp
            *     5430 167647: contig of 162218 bp in length
            *   167648 167747: gap of      100 bp
            *   167748 170305: contig of 2558 bp in length
            *   170306 170405: gap of      100 bp
            *   170406 172468: contig of 2063 bp in length
            *   172469 172568: gap of      100 bp
            *   172569 175037: contig of 2469 bp in length
            *   175038 175137: gap of      100 bp
            *   175138 177407: contig of 2270 bp in length
FEATURES             Location/Qualifiers
     source          1..177407
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /map="q22.1-31.1"
                     /clone="RP11-98F14"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..2152
```

```
     misc_feature        /note="assembly_fragment:00222"
                     2253..5329
     misc_feature        /note="assembly_fragment:01300"
                     5430..167647
     misc_feature        /note="assembly_fragment:01979"
                     167748..170305
     misc_feature        /note="assembly_fragment:02310"
                     170406..172468
     misc_feature        /note="assembly_fragment:02689"
                     172569..175037
     misc_feature        /note="assembly_fragment:03070"
                     175138..177407
     misc_feature        /note="assembly_fragment:05336"
BASE COUNT    45127 a   43957 c   44529 g   43189 t   605 others
ORIGIN

Query Match              7.3%;  Score 32.6;  DB 78;  Length 177407;
Best Local Similarity   52.6%;  Pred. No. 41;
Matches     71;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps   0;

Qy    235  tactgtgtaggcgttgatgacttcctgtgtactccgcttggagccattgagtgctg 294
           || |||||||| ||| || ||| ||||| |  |||||  ||| | | ||| |||| |||
Db  82690  TTTCTTTTGTCTCTCTGGCTCTGCCTCTCCTCTCCCTCCCCCCGCGCCTGAGCGCTG 82631

Qy    295  tttaccgtgcagcactcgaagacttccttgctgctacccgctggcagcatagtcgtttgc 354
           || ||||||| || || |||| | ||||||| ||| | || ||| |||| |||| ||||
Db  82630  TCTGACGGGAACCCCAGCAGTCTCCCTCCGCTCTCCTCCTGCAGCCTCGAGACTCCTCTGGT 82571

Qy    355  tgcctgtctgggga 369
           ||| |||| || ||
Db  82570  CTCACTGACTGCAGA 82556

RESULT 41
AL162454/c
LOCUS       AL162454    313750 bp    DNA              HTG       21-FEB-2001
DEFINITION  Homo sapiens chromosome 13 clone RP11-265C7, *** SEQUENCING IN
            PROGRESS ***, 41 unordered pieces.
ACCESSION   AL162454
VERSION     AL162454.16  GI:13121360
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 313750)
AUTHORS     Burton,J.
TITLE       Direct Submission
JOURNAL     Submitted (21-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
            On Feb 23, 2001 this sequence version replaced gi:13092262.
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----- Project Information
            Center project name: bA265C7
            -------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 295118 bases at least Q40
            Consensus quality: 301564 bases at least Q30
            Consensus quality: 305528 bases at least Q20
            Insert size: 309750; sum-of-contigs
            Insert size: 174504; 5.7% error; agarose-fp
            Quality coverage: 4.31x in Q20 bases; sum-of-contigs Quality
            coverage: 8.86x in Q20 bases; agarose-fp
            --------
            * NOTE: This is a 'working draft' sequence. It currently
```

```
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1    22725: contig of 22725 bp in length
*    22726    22825: gap of    100 bp
*    22826    27666: contig of 4841 bp in length
*    27667    27766: gap of    100 bp
*    27767    34399: contig of 6633 bp in length
*    34400    34499: gap of    100 bp
*    34500    37560: contig of 3061 bp in length
*    37561    37660: gap of    100 bp
*    37661    65584: contig of 27924 bp in length
*    65585    65684: gap of    100 bp
*    65685    69262: contig of 3578 bp in length
*    69263    69362: gap of    100 bp
*    69363    73709: contig of 4347 bp in length
*    73710    73809: gap of    100 bp
*    73810    83068: contig of 9259 bp in length
*    83069    83168: gap of    100 bp
*    83169    85224: contig of 2056 bp in length
*    85225    85324: gap of    100 bp
*    85325    87637: contig of 2313 bp in length
*    87638    87737: gap of    100 bp
*    87738    93159: contig of 5422 bp in length
*    93160    93259: gap of    100 bp
*    93260    98655: contig of 5396 bp in length
*    98656    98755: gap of    100 bp
*    98756   105031: contig of 6276 bp in length
*   105032   105131: gap of    100 bp
*   105132   111758: contig of 6627 bp in length
*   111759   111858: gap of    100 bp
*   111859   120085: contig of 8227 bp in length
*   120086   120185: gap of    100 bp
*   120186   123521: contig of 3336 bp in length
*   123522   123621: gap of    100 bp
*   123622   135166: contig of 11545 bp in length
*   135167   135266: gap of    100 bp
*   135267   139159: contig of 3893 bp in length
*   139160   139259: gap of    100 bp
*   139260   143355: contig of 4096 bp in length
*   143356   143455: gap of    100 bp
*   143456   153317: contig of 9862 bp in length
*   153318   153417: gap of    100 bp
*   153418   156147: contig of 2730 bp in length
*   156148   156247: gap of    100 bp
*   156248   159234: contig of 2987 bp in length
*   159235   159334: gap of    100 bp
*   159335   163807: contig of 4473 bp in length
*   163808   163907: gap of    100 bp
*   163908   167369: contig of 3462 bp in length
*   167370   167469: gap of    100 bp
*   167470   172739: contig of 5270 bp in length
*   172740   172839: gap of    100 bp
*   172840   186162: contig of 13323 bp in length
*   186163   186262: gap of    100 bp
*   186263   193526: contig of 7264 bp in length
*   193527   193626: gap of    100 bp
*   193627   195981: contig of 2355 bp in length
*   195982   196081: gap of    100 bp
*   196082   202543: contig of 6462 bp in length
*   202544   202643: gap of    100 bp
*   202644   216446: contig of 13803 bp in length
*   216447   216546: gap of    100 bp
*   216547   231463: contig of 14917 bp in length
*   231464   231563: gap of    100 bp
*   231564   234196: contig of 2633 bp in length
*   234197   234296: gap of    100 bp
*   234297   239413: contig of 5117 bp in length
*   239414   239513: gap of    100 bp
*   239514   242362: contig of 2849 bp in length
*   242363   242462: gap of    100 bp
*   242463   255870: contig of 13408 bp in length
*   255871   255970: gap of    100 bp
*   255971   259826: contig of 3856 bp in length
*   259827   259926: gap of    100 bp
*   259927   276107: contig of 16181 bp in length
*   276108   276207: gap of    100 bp
*   276208   293707: contig of 17500 bp in length
*   293708   293807: gap of    100 bp
*   293808   301291: contig of 7484 bp in length
*   301292   301391: gap of    100 bp
*   301392   311585: contig of 10194 bp in length
*   311586   311685: gap of    100 bp
*   311686   313750: contig of 2065 bp in length.
FEATURES             Location/Qualifiers
     source          1..313750
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-265C7"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..22725
                     /note="assembly_fragment:00276
                     fragment_chain:1"
     misc_feature    22826..27666
                     /note="assembly_fragment:00733
                     fragment_chain:1"
     misc_feature    27767..34399
                     /note="assembly_fragment:00703
                     fragment_chain:1"
     misc_feature    34500..37560
                     /note="assembly_fragment:00299
                     fragment_chain:1"
     misc_feature    37661..65584
                     /note="assembly_fragment:00236
                     fragment_chain:1"
     misc_feature    65685..69262
                     /note="assembly_fragment:01460
                     fragment_chain:1"
     misc_feature    69363..73709
                     /note="assembly_fragment:03799
                     fragment_chain:2"
     misc_feature    73810..83068
                     /note="assembly_fragment:00032
                     fragment_chain:2"
     misc_feature    83169..85224
                     /note="assembly_fragment:01349
                     fragment_chain:2"
     misc_feature    85325..87637
                     /note="assembly_fragment:01736
                     fragment_chain:2"
     misc_feature    87738..93159
                     /note="assembly_fragment:01701
                     fragment_chain:3"
     misc_feature    93260..98655
                     /note="assembly_fragment:01843
                     fragment_chain:3"
     misc_feature    98756..105031
                     /note="assembly_fragment:02264
                     fragment_chain:3"
     misc_feature    105132..111758
                     /note="assembly_fragment:00718
                     fragment_chain:4"
     misc_feature    111859..120085
                     /note="assembly_fragment:02459
                     fragment_chain:4"
     misc_feature    120186..123521
                     /note="assembly_fragment:01408
                     fragment_chain:5"
     misc_feature    123622..135166
                     /note="assembly_fragment:03466
                     fragment_chain:5"
```

misc_feature   /function="transduction of growth hormone action"
                      /evidence=experimental
                      /product="growth hormone receptor"
                      /protein_id="AAA84745.1"
                      /db_xref="GI:1142644"
                      /translation="MDLRHLLLTLVLVCANDSLSASDDVLRLPQISKCRSPELETFSC
                      YWTDGNFYNLSAPGTIQLLYMKRNDEDWKECPDYITAGENSCYFNTSYTSIWIPYCVK
                      LVNKDEVFDEKCFSVDEIVLPDPPVHLMWTLLNTSQTGIHGDIQVRWDPPPTADVQKG
                      WITLEYELQYKEVNETKWKELEPRLSTMVPLYSLKIGRDYEIRVRSRQRTSEKFGEFS
                      EILYVSFSQAGIEFVHCAEEIEFPWFLVVIFGACGLAVTVILILSKQSRLKMLIFPP
                      VPVPKIKGIDPDLLKKGKLDEVNSILASHDNYKTQLYNDDLWVEFIELDIEDPDEKNR
                      VSDTDRLLSEDHLKSHSCLGAKDDSGRASCCEPDIPETDFSASDTCDAISDIDQFKK
                      VTEKEDLLCLGRKDNDESLPSLANDTDQPRMSTRPENSQPWPPFADSIDAASPSAH
                      NQLSNQNSLRNTDFYAQVSDITPAGSVVLSPCQKSKVARARCEFCTEQNFTLDNAYFC
                      EADVKKCIAVISHEEDEPRVQAQICNEDTYFTTESLTTTGISLGASTAETPSPEVPVP
                      DYTSIHIVHSPQGLVLNATALPVPDKEFNMSCGYVSTDQLNKIMP"

misc_feature   767..838
                      /gene="ghr"
                      /note="encodes transmembrane domain"
misc_feature   839..1879
                      /gene="ghr"
                      /note="encodes intracellular domain"
      3'UTR   1883..1983
BASE COUNT   596 a   419 c   447 g   521 t
ORIGIN

Query Match   7.3%; Score 32.4; DB 8; Length 1983;
Best Local Similarity 48.0%; Pred. No. 43;
Matches 123; Conservative 0; Mismatches 131; Indels 2; Gaps 1;

Qy   97 cacggcataatggaaaacagtcctgccctttgagggtcgtccaaacatgcccctcga   156
         ||||| || || ||| | |||| || | | ||| ||| | || | || || |
Db 1787 CAAGGCCTTGTGGGAGTGGACGCATATGAATAGACACGTGTAGTCTGGGACAGGCACCTCTG 1728

Qy   157 gacgtcgg--tgaagccgtagtttgtacaagactaacaacgccaatggcgtcgatgttat 214
         || |||| | || ||| | || | || |||| || | ||| | || | || | |
Db 1727 GACTTGGGGTTTCTGCTGTTGAAGCTCCAAGACTGATGCCAGTAGTGTAAGGCTTTCTG   1668

Qy   215 cggagccattcgtgtgctgtttactgtgtagcgctgatggcctgatgacttcctttgctactccgc 274
         | ||||||| ||||| | |||| | || ||| | || | | | |||||| |||| |
Db 1667 TGGTGAAGTAAGTCCTCGTTACAGATTTGCGCGTCGAACACGTGGCCTCATCCTCTTCGT 1608

Qy   275 ttggagccattgagtcgtgttaccgtgcgcagcactcgaagacttcctttgctgtactcgc 334
         |||| |||||| ||| | ||| | || | ||||| | ||||| ||| |||| | ||
Db 1607 GGGAAATCACAGCAATACATTTTTCACATCTGCCTCACAGAGTAGGCATTGTCCAGGG   1548

Qy   335 tggcagcatagtctgt   350
         || ||||| | |||
Db 1547 TGAAGTTTGTTCTCGT 1532

RESULT 43
FRX297967
FRX297967
LOCUS   FRX297967   2263 bp   DNA   PLN   01-DEC-2000
DEFINITION Fragaria x ananassa gene encoding methionine sulfoxide reductase.
ACCESSION   AJ297967
VERSION   AJ297967.1 GI:11342532
KEYWORDS   methionine sulfoxide reductase.
SOURCE   Fragaria x ananassa
  ORGANISM Fragaria x ananassa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Rosales; Rosaceae; Fragaria.
REFERENCE   1 (bases 1 to 2263)
  AUTHORS   Pedraza-Lopez,A., Cardenas-Torres,J. and Rodriguez-Franco,A.
  TITLE   Characterization of a peptide methionine sulfoxide reductase
            isolated from strawberry fruit receptacles
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2263)
  AUTHORS   Rodriguez-Franco,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (14-FEB-1996) Rodriguez-Franco A., Universidad de

misc_feature   135267..139159
                      /note="assembly_fragment:01967
                      fragment_chain:6"
misc_feature   139260..143355
                      /note="assembly_fragment:03108
                      fragment_chain:6"
misc_feature   143356..153317
                      /note="assembly_fragment:02873
                      fragment_chain:7"
misc_feature   153418..156147
                      /note="assembly_fragment:02323
                      fragment_chain:7"
misc_feature   156248..159234
                      /note="assembly_fragment:02935
                      fragment_chain:8"
misc_feature   159335..163807
                      /note="assembly_fragment:02794

Query Match   7.3%; Score 32.6; DB 79; Length 313750;
Best Local Similarity 52.6%; Pred. No. 41;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy   235 ttactgtgtagcgcttgatgactccttgctgtactccgcttggagccattgagtgctg   294
         || || |||| | | ||||||| ||| |||| |||| | |||| |||| || |||||
Db 229989 TTTCTTTTGTCTCTCGGCTCTGCCTCTCTCCTCTCCCCCTCGCCCCCGCTGAGCGCTG 229930

Qy   295 tttaccgtgcagcactcctgctgtactcctgctgctgaccgtcagcatagtctgttgc 354
         | ||| | || |||| | |||| ||| | || || ||| || |||| || ||||
Db 229929 TCTGACGGGAACCCCCAGCAGTCTCCCCTCGCTCTCTGGGCAGCCTGCAGACTCCCTCGGT 229870

Qy   355 tgcctggtctgggga   369
         || ||| |||||
Db 229869 CTCACTGACTGCAGA 229855

RESULT 42
CLU20353/c
LOCUS   CLU20353   1983 bp   mRNA   VRT   31-DEC-1995
DEFINITION Columba livia growth hormone receptor (ghr) mRNA, complete cds.
ACCESSION   U20353
VERSION   U20353.1 GI:1142643
KEYWORDS   .
SOURCE   domestic pigeon.
  ORGANISM Columba livia
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Columbiformes; Columbidae; Columba.
REFERENCE   1 (bases 1 to 1983)
  AUTHORS   Cheng,C.H.K., Shaw,P.C., Tsim,K.W.K. and Lau,K.F.
  TITLE   Molecular cloning of the entire coding sequence of pigeon growth
            hormone receptor cDNA by PCR techniques
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1983)
  AUTHORS   Cheng,C.H.K.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-JAN-1995) Christopher H.K. Cheng, Biochemistry, The
            Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
FEATURES   Location/Qualifiers
     source   1..1983
                     /organism="Columba livia"
                     /db_xref="taxon:8932"
                     /clone="pgnGHR"
                     /sex="female"
                     /tissue_type="liver"
                     /dev_stage="adult"
     5'UTR   1..46
     gene   47..1882
                     /gene="ghr"
     misc_feature   47..766
                     /gene="ghr"
                     /note="encodes extracellular domain"
     CDS   47..1882
                     /gene="ghr"
                     /codon_start=1

```
FEATURES          Cordoba. Facultad de Ciencias, Bioquimica y Biologia Molecular, Av.
  source          S. Alberto Magno S/N, Cordoba, SPAIN, 14071
                  1..2263
                  Location/Qualifiers
                  /organism="Fragaria x ananassa"
                  /variety="Chandler"
                  /db_xref="taxon:3747"
  CDS             join(1372..1686,1813..2073)
                  /codon_start=1
                  /product="methionine sulfoxide reductase"
                  /protein_id="CAC17011.1"
                  /db_xref="GI:11342533"
                  /translation="MASSTTNNPALDLDSDTPENPGHELAQFASGCFWGSELRFQRVV
                  GVIKTEVGYSQGHVHDPNYRLVCSGTTNHSEVVRVQFDPQVCPYSDLLSVFWSRHDPT
                  TLNCQGDVGTQYRSGIYYNEEQDCLAKKSKEAKQKEFKDKRVTEILPAKRFYRAE
                  EYHQQYLEKGGGNKQSAQKGCNDPIKCYG"
  exon            <1372..1686
                  /number=1
  intron          1687..1812
                  /number=1
  exon            1813..>2073
                  /number=2
BASE COUNT      630 a    508 c    530 g    595 t
ORIGIN

Query Match         7.3%;       Score 32.4;   DB 14;   Length 2263;
Best Local Similarity 60.0%;   Pred. No. 43;
Matches 54;     Conservative 0;    Mismatches 36;   Indels 0;   Gaps 0;

Qy    206 cgatgttatcggagccattcgtcgtgcgtgttactgtgtagccgcttgatgactccttgct 265
          | ||| || |||| ||| ||  | ||||| ||||| || ||| ||| ||| |||  | |||
Db   1142 CTACGATACGGGAGTCAGGCAACTATTGTCTTCTTTGTTGTTGTGTGGACATACTTACA 1201

Qy    266 gtactccgcttggagccattgagtgctgt 295
          ||   || ||||||  |||| ||  |||||
Db   1202 ATATTCTCCTTGAACACATTGAACCGCTGT 1231

RESULT 44
AC018863/c
LOCUS       AC018863      186053 bp    DNA             HTG         12-AUG-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-20J1, WORKING DRAFT SEQUENCE,
            6 unordered pieces.
ACCESSION   AC018863
VERSION     AC018863.2  GI:9755480
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186053)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Homo sapiens clone
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 186053)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (21-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 9, 2000 this sequence version replaced gi:6624119.

            --------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ----------- Project Information --------------
            Center project name: H_NH0020J01
            --------------- Summary Statistics  ---------------
            Sequencing vector: M13; 90%
            Sequencing vector: plasmid; 10%
            Chemistry: Dye-primer ET; 90% of reads
```

```
            Chemistry: Dye-terminator Big Dye; 10% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 180709 bases at least Q40
            Consensus quality: 182200 bases at least Q30
            Consensus quality: 183266 bases at least Q20
            Insert size: 179000; agarose-fp
            Insert size: 185553; sum-of-contigs
            Quality coverage: 5.10 in Q20 bases; agarose-fp
            Quality coverage: 5.09 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *           1      6671: contig of 6671 bp in length
            *        6672      6771: gap of unknown length
            *        6772     18019: contig of 11248 bp in length
            *       18020     18119: gap of unknown length
            *       18120     33989: contig of 15870 bp in length
            *       33990     34089: gap of unknown length
            *       34090     56036: contig of 21947 bp in length
            *       56037     56136: gap of unknown length
            *       56137     97345: contig of 41209 bp in length
            *       97346     97445: gap of unknown length
            *       97446    186053: contig of 88608 bp in length.
FEATURES             Location/Qualifiers
  source     1..186053
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-20J1"
  misc_feature       1..6671
                     /note="assembly_name:Contig11"
  misc_feature       6772..18019
                     /note="assembly_name:Contig12"
  misc_feature       18120..33989
                     /note="assembly_name:Contig13"
                     clone_end:SP6
                     vector_side:right"
  misc_feature       34090..56036
                     /note="assembly_name:Contig14"
                     clone_end:T7
                     vector_side:right"
  misc_feature       56137..97345
                     /note="assembly_name:Contig15"
  misc_feature       97446..186053
                     /note="assembly_name:Contig16"
BASE COUNT      58625 a    33523 c    33285 g    60118 t    502 others
ORIGIN

Query Match               7.3%;       Score 32.4;   DB 65;   Length 186053;
Best Local Similarity   58.2%;    Pred. No. 47;
Matches 57;   Conservative 0;   Mismatches 41;   Indels 0;   Gaps 0;

Qy    218 agccattcgtcgtgcgtgttactgtgtagccgcttgatgactccttgctgtactccgcttg 277
          | ||||| |  |||  |||  |  ||| || ||||  ||  |||| ||| || ||| |||||
Db   78936 AACCATGCCCCGTGTTCTGGGATTGGAGTTAGCCTTGAGGATTTCCTGGCTTTACTCCGGCTG 78877

Qy    278 gagccattgagtgctgttaccgtgcagcactcgaag 315
          |  | |||| |  | ||| ||| || | || | | ||
Db   78876 CAGTTGTTTTCAGTATTAATTTGCCTCTCCCTCAG 78839

RESULT 45
AC013799/c
LOCUS       AC013799      190650 bp    DNA             HTG         16-MAR-2000
DEFINITION  Homo sapiens clone RP11-20J1, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
```

```
ACCESSION   AC013799
VERSION     AC013799.3  GI:7107979
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 190650)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-20J1
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 190650)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,M., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,Y.W.J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6563543.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: L3913
            Center clone name: 20_J_1
            --------------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 168113 bases at least Q40
            Consensus quality: 180987 bases at least Q30
            Consensus quality: 186212 bases at least Q20
            Insert size: 188000; agarose-fp
            Insert size: 189650; sum-of-contigs
            Quality coverage: 5.1 in Q20 bases; agarose-fp
            Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1  1030: contig of 1030 bp in length
            *   1031  1130: gap of    100 bp
            *   1131  2643: contig of 1513 bp in length
            *   2644  2743: gap of    100 bp
            *   2744  4083: contig of 1340 bp in length
            *   4084  4183: gap of    100 bp
            *   4184  5255: contig of 1072 bp in length
            *   5256  5355: gap of    100 bp
            *   5356  6639: contig of 1284 bp in length
            *   6640  6739: gap of    100 bp
            *   6740 11183: contig of 4444 bp in length
            *  11184 11283: gap of    100 bp
```

```
            *  11284 31827: contig of 20544 bp in length
            *  31828 31927: gap of    100 bp
            *  31928 52477: contig of 20550 bp in length
            *  52478 52577: gap of    100 bp
            *  52578 85160: contig of 32583 bp in length
            *  85161 85260: gap of    100 bp
            *  85261 128272: contig of 43012 bp in length
            *  128273 128372: gap of   100 bp
            *  128373 190650: contig of 62278 bp in length.
FEATURES             Location/Qualifiers
     source          1..190650
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-20J1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1030
                     /note="assembly_fragment"
     misc_feature    1131..2643
                     /note="assembly_fragment"
     misc_feature    2744..4083
                     /note="assembly_fragment"
     misc_feature    4184..5255
                     /note="assembly_fragment"
     misc_feature    5356..6639
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    6740..11183
                     /note="assembly_fragment"
     misc_feature    11284..31827
                     /note="assembly_fragment"
     misc_feature    31928..52477
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    52578..85160
                     /note="assembly_fragment"
     misc_feature    85261..128272
                     /note="assembly_fragment"
     misc_feature    128373..190650
                     /note="assembly_fragment"
BASE COUNT     60534 a   34407 c   34001 g   60701 t    1007 others
ORIGIN

Query Match     7.3%;  Score 32.4;  DB 63;  Length 190650;
Best Local Similarity  58.2%;  Pred. No. 47;
Matches   57;  Conservative   0;  Mismatches  41;  Indels   0;  Gaps   0;

Qy    218 agccattcgtgtgctgttactgtgtagcgcttgatgacttccttgctgtgtactccgcttg 277
          ||||| |||| ||||| |||||||||||  |||||| ||||| |||||||||| || ||
Db 111444 AACCATGCCCCTGTTCTTGGGATTGTAGCCTTTGAGGATTTCCTGGCTTTACTCCGGCTG 111385

Qy    278 gagccattgagtgctgttaccgtgcagcactcgaag 315
          || ||| |||| ||||| |||||| ||| ||| || ||
Db 111384 CAGTTGTTTCTCAGTATTAATTTGCCTCTCCCTCAG 111347

Search completed: September  7, 2001, 17:04:05
Job time: 7509 sec
```

THIS PAGE BLANK (USPTO)